



# Etude contrastive de la prosodie audio-visuelle des affects sociaux en chinois mandarin vs.français : vers une application pour l'apprentissage de la langue étrangère ou seconde

Yan Lu

## ► To cite this version:

Yan Lu. Etude contrastive de la prosodie audio-visuelle des affects sociaux en chinois mandarin vs.français : vers une application pour l'apprentissage de la langue étrangère ou seconde. Linguistique. Université Grenoble Alpes, 2015. Français. NNT : 2015GREAL001 . tel-01227267

**HAL Id: tel-01227267**

**<https://theses.hal.science/tel-01227267>**

Submitted on 10 Nov 2015

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## THÈSE

Pour obtenir le grade de

## DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Spécialité : Sciences du langage Spécialité Linguistique,  
Sociolinguistique et Acquisition du langage

Arrêté ministériel : 7 août 2006

Présentée par

**Yan LU**

Thèse dirigée par **Didier DEMOLIN** et **Véronique AUBERGE**

préparée au sein du **Laboratoire GIPSA-lab**  
dans l'**École Doctorale Langues, Littératures et Sciences Humaines**

# Etude contrastive de la prosodie audio-visuelle des affects sociaux chinois mandarin vs. français vers une application pour l'apprentissage de la langue étrangère ou seconde

Thèse soutenue publiquement le **22 janvier 2015**,  
devant le jury composé de :

**M. Jürgen TROUVAIN**

Professeur à l'Université des Saarlandes, Rapporteur

**Mme. Anna DE MEO**

Professeur à l'Université de Naples "L'Orientale", Rapporteur

**M. Laurent BESACIER**

Professeur à l'Université Joseph Fourier – Grenoble I, Président

**M. Albert RILLIARD**

CR-CNRS au LIMSI, UPR CNRS 3251, Examineur

**M. Wentao GU**

Professeur à l'Université Normale de Nanjing, Examineur

**M. Didier DEMOLIN**

Professeur à l'Université Sorbonne Nouvelle – Paris 3, Directeur de thèse

**Mme. Véronique AUBERGE**

CR-CNRS au LIG, UMR CNRS 5217, co-directrice de thèse

*Université Joseph Fourier / Université Pierre Mendès France /  
Université Stendhal / Université de Savoie / Grenoble INP*





# Remerciements

Tout d'abord, je voudrais remercier Véro, ma directrice, de m'avoir proposé un sujet super intéressant et passionnant, de m'avoir fait une confiance incroyable et de m'avoir accordé une grande liberté. Merci de m'avoir accueillie et de m'avoir encadrée scientifiquement pendant ces quatre années.

Je voudrais aussi remercier Didier, mon directeur de thèse officiel, pour son encouragement et sa sollicitude. Merci beaucoup d'avoir toujours été présent pour m'aider à régler les problèmes administratifs et ainsi me permettre de travailler sereinement.

Merci au Pr. Jürgen Trouvain et au Pr. Anna De Meo d'avoir accepté de rapporter cette thèse. Merci à tous les membres de ce jury de vous être intéressés à mon travail et d'avoir accepté de relire cette thèse. Je remercie en particulier Laurent Besacier d'avoir accepté d'être le président de ce jury.

Un grand merci à Albert. Merci d'avoir toujours répondu présent pour me sauver des difficultés statistiques et techniques, et ce malgré la distance géographique. J'ai eu souvent des remords de te déranger avec mes petites choses, mais chaque fois que j'étais coincée dans mon travail, c'est à toi que je pensais... Je te remercie pour ta patience, tes conseils, ton soutien et surtout pour tout le temps consacré à mes expériences, à mes données et à la relecture de ma thèse. Sans toi je n'aurais pas pu finir ma thèse.

Merci au département Parole et Cognition du laboratoire GIPSA-lab et plus personnellement à Denis Beautemps et Nathalie Vallée de m'avoir donné la possibilité de terminer cette thèse dans les meilleures conditions.

Merci à Silvain Gerber pour ses connaissances et ses conseils en statistiques. Ton aide m'a permis de résoudre de nombreuses difficultés dans l'analyse de données.

Je tiens à remercier tous les sujets de mes tests perceptifs d'avoir accepté de passer du temps pour ma recherche, que ce soit pour me donner un coup main par amitié ou par un intérêt particulier pour mon sujet. Sans eux mes recherches n'auraient pas été possibles. Parmi ces 314 juges, mis à part mes amis et mes collègues, il y a beaucoup de



personnes qui ne me connaissaient pas, mais qui ont la gentillesse de dire « oui » à la vue d'un *Moi* quasiment désespérée par le recrutement des sujets. Certaines personnes ont même été torturées plusieurs fois. Merci de ne pas m'en avoir voulu. ☺

Manon, un grand merci pour tout ! Dix pages suffiraient à peine pour énumérer tout ce que tu as fait pour moi. Merci pour ta présence quand je travaillais dans la solitude ; merci pour tes consolations quand j'ai perdu le morale ; merci pour tes propositions quand j'étais confuse ; merci pour ton empathie quand j'étais triste ; merci d'être contente quand j'étais contente ; merci pour ta formation de français, à Praat et à « Word » ; merci d'avoir passé ton temps à relire mes chapitres alors qu'il aurait du être consacré à ton propre travail ; merci pour... tant de choses ! Si je pouvais faire autant que tu en as fait pour moi ! Je n'oublierai jamais toutes les journées et tous les soirs où l'on a travaillé côte à côte dans notre coin au labo ; les happy hours qu'on a passé avec Rita dans un bar ; les petits secrets de filles qu'on a partagés l'une avec l'autre ; tes sourires, ton humour, tes caprices, tes robes chouettes et tes bottines bleues parfaites. ☺

Merci à mon équipe de relecteurs d'avoir relu les chapitres et d'avoir corrigé mes phrases incompréhensibles et structurées à la chinoise. Yoann, merci pour tes précieux conseils sur la présentation de la langue chinoise ; Lucie et Avril, merci de m'avoir donné la priorité dans votre emploi du temps et d'avoir toléré mes caprices causés par le stress. Michèle, merci pour tes corrections minutieuses et ta volonté constante de m'aider. Merci à André et à Monsieur et Madame Revon de m'avoir donné un coup de main. Merci à Nathalie H d'avoir accepté de corriger mon anglais maladroit.

Merci à ceux du site Stendhal du laboratoire, permanents, doctorants, masters ou de passage au cours de ces quatre années, qui m'ont apporté leur soutien psychologique et avec qui j'ai eu des échanges scientifiques et humains : Carole, Coriandre, Christophe, Anne, Rita, Hien, Diane, Silvia, Katja et les autres. Merci pour votre aide et vos conseils, vos sourires et vos petits mots d'encouragement et de souhait, pour vos éclats de rire à la cafète. Tout ça m'a permis de me sentir encore vivante sous la pression et le stress de travail.

Merci aussi à ceux externes du laboratoire, surtout les membres de l'équipe GETALP du LIG, qui m'ont souvent aidé dans mon travail : Solange, Fred, Yuko et les autres. Un grand merci à Mme. Monica Masperi et son équipe de professeurs de chinois à LANSAD

de m'avoir aidé pour le recrutement des sujets. Je remercie également à Isabelle Rousset pour son soutien psychologique et pour ses précieux conseils pour affronter la fin de la thèse.

Merci à tous mes amis chinois rencontrés à Grenoble, merci de m'avoir accompagnée au cours de ces quatre années. Votre amitié sera un des plus beaux souvenirs dans ma vie.

Merci à ma famille qui m'a soutenue depuis le jour où j'ai décidé de venir en France, surtout à mes parents, qui ont été obligés de s'occuper de mon fils à ma place et qui étaient, heureusement, très partants pour le faire. Sans eux, je n'aurais pas pu venir ni travailler en toute tranquillité. Merci à mon mari de m'avoir soutenue comme toujours. Merci à mon petit garçon, Ruirui, merci de ne pas en vouloir à sa maman qui l'a abandonné pendant des années et merci pour son immense et tendre amour qui fut la seule motivation de sa maman à faire tout son possible pour terminer au plus vite et au mieux cette thèse.

Merci à tous ceux que j'oublie.

# Résumé

Se distinguant des expressions émotionnelles qui sont innées et déclenchées par un contrôle involontaire du locuteur au sein d'une communication face-à-face, les affects sociaux émergent plutôt de manière volontaire et intentionnelle, et sont largement véhiculés par la prosodie audio-visuelle. Ils mettent en circulation, entre les interactants, des informations sur la dynamique du dialogue, la situation d'énonciation et leur relation sociale. Ces spécificités culturelles et linguistiques de la prosodie socio-affective dans la communication orale constituent une difficulté, même un risque de malentendu, pour les apprenants en langue étrangère (LE) et en langue seconde (L2).

La présente thèse se consacre à des études intra- et interculturelles sur la perception de la prosodie de 19 affects sociaux en chinois mandarin et en français, ainsi que sur leurs représentations cognitives. Son but applicatif vise l'apprentissage de la prosodie des affects sociaux en chinois mandarin et en français LE ou L2.

Le premier travail de la thèse consiste en la construction d'un large corpus audio-visuel des affects sociaux chinois. 152 énoncés variés dans leur longueur, leur morpho-syntaxe et leur représentation tonale sont respectivement produits dans les 19 affects sociaux. Sur la base de ce corpus, sont examinées l'identification et les confusions perceptives de ces affects sociaux chinois par des natifs, des français et des vietnamiens (comme groupe de référence), ainsi que l'effet du ton lexical sur l'identification auditive des sujets non natifs. Les résultats montrent que la majorité des affects sociaux chinois est perçue de manière similaire par les sujets natifs et les sujets non natifs, cependant certains décalages perceptifs sont également observés. Les tons chinois engendrent des problèmes perceptifs des affects sociaux autant pour les vietnamiens (d'une langue tonale) que pour les français (d'une langue non tonale). En parallèle, une analyse acoustique permet de mettre en évidence les caractéristiques principales de la prosodie socio-affective en chinois et d'étayer les résultats perceptifs.

Ensuite, une étude sur les distances conceptuelles d'une part, et psycho-acoustiques d'autre part, entre les affects sociaux est menée auprès de sujets chinois et de sujets français. Les résultats montrent que la plupart des connaissances sur les affects sociaux sont partagées par les sujets, quels que soient leur langue maternelle, leur genre ou la manière de présenter les affects sociaux (concept ou entrée acoustique).

Enfin, le dernier chapitre de la thèse est consacré à une étude contrastive sur la perception multimodale des affects sociaux en chinois et en français LE ou L2. Il est constaté que la reconnaissance des affects sociaux est étroitement liée aux expressions elles-mêmes et à la modalité de présentation de ces expressions. Le degré d'acquisition de la langue cible du sujet n'a pas d'impact significatif sur la reconnaissance, dans le cadre restreint des niveaux étudiés.

**Mots clés :** *attitudes, affects sociaux, prosodie audio-visuelle, perception interculturelle de la prosodie expressive, langue tonale, chinois, français, apprentissage, langue étrangère, L2.*

# Abstract

In human face-to-face interaction, social affects are expressed within the voluntarily controlled audio-visual prosody. As such, they should be distinguished from emotional expressions, triggered in speakers by innate and involuntary controls. This distinction in the nature of these two kinds of affects is important to understand the specific role of social affects in the realization of speech acts, as well as in signaling the social context of communication and the social relationship between interlocutors. Prosody is an important vector of such social affects, and its cross-linguistic variability is a challenge for language description as well as for foreign language teaching. Cultural and linguistic specificities of socio-affective prosody in oral communication could raise some difficulties during interactions, even risks of misunderstanding, for foreign and second language learners.

This thesis is dedicated to the intra- and intercultural study of the prosody of 19 social affects in Mandarin Chinese and in French: from their perception and acoustic characterization to their cognitive representations. It also addresses the multimodal perception of Chinese and French socio-affective prosodic variations by foreign and second language learners.

The first task of this thesis concerns the construction of an audio-visual corpus of Chinese social affects. 152 utterances with variations of length, tone type and location, and syntactic structures of utterances, are produced to express 19 social affects. This corpus allows the examination of the identification and perceptual confusions of these Chinese social affects by native and non-native listeners. It also allows the investigation of the effect of tones on non-native subjects' identification of social affects. Experimental results reveal that most social affects are perceived comparably by native and non-native subjects. Some differences are also observed and discussed. The presence of lexical tones leads to some perceptual mismatches for both Vietnamese and French listeners (having or not a tonal language as L1). An acoustic analysis investigates the production side of prosodic socio-affects in Mandarin Chinese, and highlights their main acoustic patterns. It supports the description of the perceptual results obtained on the same expressions.

Then, a study on the conceptual and psycho-acoustic distances between social affects is carried out with Chinese and French subjects. The results indicate that all subjects share to a large extent the same cognitive representation of these 19 social affects, regardless of their native language, gender or the modality used to present the social affective meanings (concept or acoustic realization).

Finally, the last chapter of this thesis observes the differences in the perception of 11 Chinese social affects expressed in different modalities (audio-only, video-only and audio-visual) for French learners and native subjects. It also describes the perception of the same socio-affects in French, as perceived by Chinese learners and native subjects. According to the results, the identification of socio-affective expressions mostly depends on their affective values and on the modality of presentation. The learner's language level (beginner or intermediate) does not have a significant effect on the identification score.

**Key words:** *attitudes, social affects, audio-visual prosody, cross-language perception of expressive prosody, tonal language, Mandarin Chinese, French, foreign language and L2 learning.*

# Table des matières

<b>REMERCIEMENTS</b>	<b>I</b>
<b>RÉSUMÉ</b>	<b>IV</b>
<b>ABSTRACT</b>	<b>V</b>
<b>TABLE DES MATIÈRES</b>	<b>VI</b>
<b>LISTE DES FIGURES</b>	<b>XII</b>
<b>LISTE DES TABLEAUX</b>	<b>XXI</b>
<b>INTRODUCTION</b>	<b>1</b>
<b>CHAPITRE 1. ANALYSE DE L'ÉTAT DE L'ART</b>	<b>8</b>
<b>1.1. Introduction</b>	<b>8</b>
<b>1.2. Quelques notions sur les affects</b>	<b>8</b>
1.2.1. Affects en psychologie	8
1.2.1.1. Emotions en psychologie	9
1.2.1.1.1. Perspective évolutionniste	9
1.2.1.1.2. Perspective physiologique	10
1.2.1.1.3. Perspective cognitive	11
1.2.1.1.4. Perspective socio-constructiviste	12
1.2.1.2. Attitudes en psychologie sociale	13
1.2.1.2.1. Notion de l'attitude	13
1.2.1.2.2. Modèle unidimensionnel classique	14
1.2.1.2.3. Modèle multidimensionnel ou modèle tripartite	14
1.2.2. Affects en linguistique et en phonostylistique	16
1.2.2.1. Distinction entre l'émotion et l'attitude	17
1.2.2.2. Quelques points de vue sur la catégorisation des attitudes	19

1.2.2.3.	Des attitudes aux affects sociaux	21
1.2.2.4.	Etiquettes données aux émotions et aux affects sociaux	22
<b>1.3.</b>	<b>Survol des théories sur la prosodie</b>	<b>24</b>
1.3.1.	Notion générale de la prosodie	25
1.3.2.	Paramètres de la prosodie	26
1.3.3.	Fonctions plurielles de la prosodie	29
1.3.3.1.	Du point de vue de la linguisticité	29
1.3.3.2.	Du point de vue fonctionnel	33
1.3.4.	Prosodie dans le système de communication parlée: le modèle C-Clone	34
1.3.5.	Quelques premières études de la didactique de la prosodie attitudinale	36
1.3.6.	Enseignement de la prosodie du Français Langue Etrangère en Chine	39
<b>1.4.</b>	<b>Chinois mandarin et sa prosodie</b>	<b>41</b>
1.4.1.	Présentation du chinois mandarin	41
1.4.1.1.	Chinois mandarin et les langues chinoises	41
1.4.1.2.	Structure syllabique du chinois mandarin	43
1.4.1.3.	Tons lexicaux du chinois mandarin	44
1.4.2.	Prosodie du chinois mandarin	45
1.4.2.1.	Fonctions linguistiques de la prosodie du chinois mandarin	46
1.4.2.2.	Fonction expressive de la prosodie du chinois mandarin : expressions émotionnelles et attitudinales en chinois	50
<b>1.5.</b>	<b>Conclusion du chapitre</b>	<b>54</b>
<b>CHAPITRE 2.</b>	<b>ETUDE INTRA-CULTURELLE DE LA PROSODIE DES AFFECTS SOCIAUX CHINOIS</b>	<b>57</b>
<b>2.1.</b>	<b>Introduction</b>	<b>57</b>
<b>2.2.</b>	<b>Construction du corpus audio-visuel des affects sociaux en chinois mandarin</b>	<b>58</b>
2.2.1.	Choix des affects sociaux	58
2.2.2.	Composition du corpus	63
2.2.3.	Enregistrement du corpus	65
2.2.3.1.	Locuteur	65
2.2.3.2.	Enregistrement	66
<b>2.3.</b>	<b>Perception intra-culturelle de la prosodie des affects sociaux chinois</b>	<b>67</b>
2.3.1.	Introduction	67
2.3.2.	Sous-corpus	67

2.3.3.	Sujets	68
2.3.4.	Protocole expérimental	69
2.3.5.	Analyse des résultats	69
2.3.5.1.	Facteurs influant la reconnaissance des affects sociaux	69
2.3.5.2.	Taux d'identification des affects sociaux	70
2.3.5.3.	Effet de la longueur	71
2.3.5.4.	Effet du genre	72
2.3.5.5.	Confusion perçue des affects sociaux	73
2.3.5.6.	Classement des affects sociaux	74
2.3.6.	Conclusion	78
<b>2.4.</b>	<b>Analyse acoustique des affects sociaux chinois</b>	<b>79</b>
2.4.1.	Introduction	79
2.4.2.	Sous-corpus	80
2.4.3.	Paramètres prosodiques et extraction des valeurs acoustiques	80
2.4.4.	Analyse statistique et résultats	82
2.4.4.1.	Classement des affects sociaux	82
2.4.4.2.	Caractéristiques acoustiques des affects sociaux	87
2.4.5.	Analyse des contours intonatifs des affects sociaux	90
2.4.5.1.	Méthodologie	90
2.4.5.2.	Classement des affects sociaux	92
2.4.5.3.	Confusion perçue des affects sociaux	108
2.4.6.	Conclusion	114
<b>2.5.</b>	<b>Perception multimodale des affects sociaux chinois</b>	<b>116</b>
2.5.1.	Introduction	116
2.5.2.	Modalité de l'expression des affects sociaux	117
2.5.3.	Sous-corpus	117
2.5.3.1.	Sélection des affects sociaux	117
2.5.3.2.	Composition du sous-corpus	119
2.5.4.	Sujets	120
2.5.5.	Protocole expérimental	121
2.5.6.	Analyse des résultats	121
2.5.6.1.	Facteurs influant la reconnaissance des affects sociaux	121
2.5.6.2.	Effet de la modalité et de l'ordre de présentation des modalités	123
2.5.6.3.	Confusion perçue des affects sociaux	126
2.5.7.	Conclusion	132
<b>2.6.</b>	<b>Conclusion du chapitre</b>	<b>134</b>

## **CHAPITRE 3. ETUDE INTERCULTURELLE DE LA PROSODIE SOCIO-AFFECTIVE EN CHINOIS ET DES CONCEPTS DES AFFECTS SOCIAUX 138**

<b>3.1. Introduction</b>	<b>138</b>
<b>3.2. Perception interculturelle des affects sociaux chinois</b>	<b>140</b>
3.2.1. Introduction	140
3.2.2. Sous-corpus	140
3.2.3. Sujets	141
3.2.1. Protocole expérimental	141
3.2.2. Analyse des résultats	142
3.2.2.1. Facteurs influant la reconnaissance des affects sociaux	142
3.2.2.2. Taux d'identification des affects sociaux	144
3.2.2.3. Effet de la langue maternelle des sujets	144
3.2.2.4. Effet du ton	145
3.2.2.5. Confusion perçue des affects sociaux	148
3.2.2.6. Classement des affects sociaux	151
3.2.3. Conclusion	154
<b>3.3. Perception conceptuelle et psycho-acoustique des affects sociaux</b>	<b>155</b>
3.3.1. Introduction	155
3.3.2. Concept des affects sociaux	156
3.3.3. Méthodologie	157
3.3.4. Mesure de la proximité conceptuelle des affects sociaux du chinois et du français	158
3.3.4.1. Introduction	158
3.3.4.2. Sujets	158
3.3.4.3. Protocole expérimental	158
3.3.4.4. Analyse des résultats	159
3.3.4.4.1. Distribution des concepts d'affects sociaux	159
3.3.4.4.2. Classement des concepts d'affects sociaux	162
3.3.4.4.3. Proportion de connaissances partagées entre les sujets	164
3.3.4.5. Conclusion	165
3.3.5. Mesure de la proximité entre le concept et la réalisation acoustique des affects sociaux chinois	167
3.3.5.1. Introduction	167
3.3.5.2. Mesure de la distance psycho-acoustique des affects sociaux chinois	167
3.3.5.2.1. Stimuli	167
3.3.5.2.2. Sujets	167
3.3.5.2.3. Protocole expérimental	167
3.3.5.3. Comparaison entre la proximité conceptuelle et la proximité acoustique des affects sociaux chinois	168



3.3.5.3.1.	Distribution des concepts et des réalisations acoustiques des affects sociaux	168
3.3.5.3.2.	Classement des affects sociaux	170
3.3.5.3.3.	Proportion de connaissances partagées entre les sujets	173
3.3.5.4.	Conclusion	174
<b>3.4.</b>	<b>Conclusion du chapitre</b>	<b>176</b>
<b>CHAPITRE 4.</b>	<b>ETUDE DE L'APPRENTISSAGE DE LA PROSODIE SOCIO-AFFECTIVE DU FRANÇAIS ET DU CHINOIS MANDARIN</b>	<b>179</b>
<b>4.1.</b>	<b>Introduction</b>	<b>179</b>
<b>4.2.</b>	<b>Hypothèse et objectif</b>	<b>179</b>
<b>4.3.</b>	<b>Construction du corpus audio-visuel des affects sociaux français</b>	<b>181</b>
4.3.1.	Sélection des affects sociaux	181
4.3.2.	Composition du corpus	181
4.3.3.	Enregistrement du corpus	182
4.3.3.1.	Locuteur	182
4.3.3.2.	Enregistrement	182
<b>4.4.</b>	<b>Perception multimodale des affects sociaux français par des sujets natifs</b>	<b>182</b>
4.4.1.	Introduction	182
4.4.2.	Sujets	183
4.4.3.	Protocole expérimental	183
4.4.4.	Analyse des résultats	183
4.4.4.1.	Facteurs influant la reconnaissance des affects sociaux	183
4.4.4.2.	Effet de la modalité et de l'ordre de présentation des modalités	185
4.4.4.3.	Confusion perçue des affects sociaux	187
4.4.5.	Conclusion	193
<b>4.5.</b>	<b>Perception multimodale des affects sociaux français par des apprenants chinois</b>	<b>194</b>
4.5.1.	Introduction	194
4.5.2.	Sujets	195
4.5.3.	Sous-corpus	195
4.5.4.	Protocole expérimental	196
4.5.5.	Analyse des résultats	196
4.5.5.1.	Facteurs influant la reconnaissance des affects sociaux	196
4.5.5.2.	Effet de la modalité et de l'ordre de présentation des modalités	197
4.5.5.3.	Confusion perçue des affects sociaux	199

4.5.5.4.	Comparaison entre les apprenants chinois et les sujets natifs	205
4.5.6.	Conclusion	208
<b>4.6.</b>	<b>Perception multimodale des affects sociaux chinois par des apprenants français</b>	<b>209</b>
4.6.1.	Introduction	209
4.6.2.	Sous-corpus	210
4.6.3.	Sujets	210
4.6.4.	Protocole expérimental	211
4.6.5.	Analyse des résultats	211
4.6.5.1.	Facteurs influant la reconnaissance des affects sociaux	211
4.6.5.2.	Effet de la modalité	212
4.6.5.3.	Confusion perçue des affects sociaux	213
4.6.5.4.	Comparaison entre les apprenants français et les sujets natifs	220
4.6.6.	Conclusion	222
<b>4.7.</b>	<b>Conclusion du chapitre</b>	<b>224</b>
	<b>CONCLUSIONS ET PERSPECTIVES</b>	<b>230</b>
	<b>PUBLICATIONS</b>	<b>242</b>
	<b>BIBLIOGRAPHIE</b>	<b>243</b>
	<b>ANNEXE 1: LES ÉNONCÉS DU CORPUS D'AFFECTS SOCIAUX CHINOIS</b>	<b>258</b>
	<b>ANNEXE 2 : INTERFACES DES TESTS PERCEPTIFS EN CHINOIS MANDARIN ET EN FRANÇAIS</b>	<b>263</b>
	<b>ANNEXE 3 : CONTOURS INTONATIFS DES AFFECTS SOCIAUX CHINOIS</b>	<b>266</b>
	<b>ANNEXE 4 : MATRICES DE CONFUSION DES TESTS PERCEPTIFS</b>	<b>273</b>

# Liste des figures

Figure 1: Représentation du modèle tripartite révisé de Zanna et Rempel (Vallerand & Lafrenaye, 2006, p. 244).	16
Figure 2 : Le modèle fonctionnel de la prosodie proposé par Aubergé (2002a, 2003)	34
Figure 3 : Représentation du système de communication réparti sur les multi-agents en coopération interactive (Aubergé & Rilliard, 2006)	35
Figure 4 : Carte de la distribution géographique des langues chinoises (Li & Tompson, 1981).	43
Figure 5 : Représentation schématique des quatre tons du chinois mandarin proposée par Chao (1968) ( <a href="http://www.chine-culture.com/chinois/pinyin.php">http://www.chine-culture.com/chinois/pinyin.php</a> )	45
Figure 6 : Contours de F0 des quatre tons du chinois mandarin présentés dans le contexte segmental de/ma/, prononcé isolément par une locutrice native (Moore & Jongman, 1997).	46
Figure 7 : Paradigme des études sur les attitudes selon le modèle d'Aubergé (Shochi, 2008, p. 43)	56
Figure 8 : Taux d'identification des 19 affects sociaux chinois pour 30 sujets natifs. Les affects sociaux sont présentés en abscisse et les scores de reconnaissance en ordonnée. La ligne rouge pointillée indique le seuil du hasard (5,3%).	71
Figure 9 : Effet de la longueur sur les taux d'identification de 19 affects sociaux chinois pour 30 sujets natifs. Les affects sociaux sont présentés en abscisse et les scores de reconnaissance en ordonnée. La ligne rouge pointillée indique le seuil du hasard (5,3%).	72
Figure 10 : Effet du genre sur les taux d'identification des 19 affects sociaux chinois pour 30 sujets natifs. Les affects sociaux sont présentés en abscisse et les scores de reconnaissance en ordonnée. Dans la légende, la lettre F indique les femmes et M les hommes. La ligne rouge pointillée indique le seuil du hasard (5,3%).	73
Figure 11 : Confusions perçues des 19 affects sociaux chinois par les sujets natifs : les pourcentages d'identification sont indiqués dans les ovales et les taux de confusions sur les flèches. Seules les confusions qui affichent des scores supérieurs à deux fois le seuil du hasard sont rapportées.	74
Figure 12 : Classification hiérarchique des 19 affects sociaux chinois pour les sujets natifs.	75
Figure 13 : Classification des affects sociaux en anglais (a), français (b), japonais (c) et vietnamien (d). Signification des étiquettes : AD (admiration), AU-AUT-CO (autorité), DC-DEC (déclaration), DO-DOU (doute), EV-EVI (évidence), EX-SU (exclamation de surprise), EXp (exclamation de surprise positive), EXo (exclamation de surprise neutre), EXn (exclamation de surprise négative), IN-FAM (intimité-familiarité), IR-IRR (irritation), KYO (kyoshuku), ME-MEP-AL-AR (mépris-arrogance),	

MAT(maternel), PO-POL (politesse), QS (question), SE (séduction), SI-SA-SAR (ironie sarcastique), SIN (sincérité) (Shochi, 2008; Mac, 2012). 85

Figure 14 : Classification hiérarchique des 19 affects sociaux à partir de leurs indices prosodiques. La ligne rouge indique les groupes obtenus selon la réduction de variance présentée dans le graphique en haut à droite. 85

Figure 15 : Représentation des deux dimensions principales de l'ACP, qui expliquent 82% de la variance et sur lesquelles les affects sociaux sont projetés. Les chiffres en-dessous des étiquettes indiquent à quel groupe cet affect social appartient. 86

Figure 16 : Valeurs moyennes des indices prosodiques pour les 5 groupes d'affects sociaux. Les chiffres sur l'abscisse représentent les groupes obtenus dans la classification hiérarchique, et ceux au-dessus des barres les valeurs moyennes d'un paramètre acoustique donné. 89

Figure 17 : Dendrogrammes montrant les distances entre les affects sociaux exprimés avec un mot monosyllabique basées sur les coefficients de corrélation pondérés (en haut) et sur la différence moyenne quadratique du registre de F0 (en bas) entre les contours intonatifs des affects sociaux. 93

Figure 18 : Comparaison de la courbe intonative du *mépris* avec celle de l'exclamation de *surprise neutre*, de la *question*, de la *déclaration* et de la *politesse*. Les valeurs de F0 des contours sont représentées en demi-ton et normalisés à 10 points par voyelle.  $r$  ( $-1 < r < 1$ ): coefficient de corrélation pondéré entre deux courbes intonatives;  $Z$ : score de la transformation de Fisher;  $rms$  ( $rms > 0$ ): différence moyenne quadratique entre deux courbes intonatives. 95

Figure 19 : Dendrogrammes montrant les distances entre les affects sociaux exprimés avec un mot dissyllabique basées sur les coefficients de corrélation pondérés (en haut) et sur la différence moyenne quadratique du registre de F0 (en bas) entre les contours intonatifs des affects sociaux. 96

Figure 20 : Comparaison de la courbe intonative de la *séduction* et de l'*irritation* avec celle de l'exclamation de *surprise neutre*, de la *question*, de la *déclaration* et de la *politesse*. Les valeurs de F0 des contours sont représentées en demi-ton et normalisés à 10 points par voyelle.  $r$  ( $-1 < r < 1$ ): coefficient de corrélation pondéré entre deux courbes intonatives;  $Z$ : score de la transformation de Fisher;  $rms$  ( $rms > 0$ ): différence moyenne quadratique entre deux courbes intonatives. 97

Figure 21 : Comparaison de la courbe intonative du *maternel*, du *doute* et du *mépris* avec celle de la *déclaration* de la *politesse*. Les valeurs de F0 des contours sont représentées en demi-ton et normalisés à 10 points par voyelle.  $r$  ( $-1 < r < 1$ ): coefficient de corrélation pondéré entre deux courbes intonatives;  $Z$ : score de la transformation de Fisher;  $rms$  ( $rms > 0$ ): différence moyenne quadratique entre deux courbes intonatives. 98

Figure 22 : Dendrogrammes montrant les distances entre les affects sociaux exprimés avec une phrase de quatre syllabes basées sur les coefficients de corrélation pondérés (en haut) et sur la différence moyenne quadratique du registre de F0 (en bas) entre les contours intonatifs des affects sociaux. 99

Figure 23 : Les courbes intonatives du *doute*, du *maternel*, de la *question*, de l'exclamation de *surprise positive* et de *surprise négative*, de la *déclaration* et de la *politesse*. Les valeurs de F0 des contours sont représentées en demi-ton et normalisés à 10 points par voyelle.  $r$  ( $-1 < r < 1$ ): coefficient de corrélation pondéré entre deux courbes intonatives;  $Z$ : score de la transformation de Fisher ;  $rms$  ( $rms > 0$ ): différence moyenne quadratique entre deux courbes intonatives. 101

Figure 24 : Comparaison de la courbe intonative de l'*intimité* avec celle de la *déclaration*, de l'*irritation*, de l'exclamation de *surprise neutre* et de *surprise négative*. Les valeurs de F0 des contours sont représentées en demi-ton et normalisés à 10 points par voyelle.  $r$  ( $-1 < r < 1$ ): coefficient de corrélation pondéré entre deux courbes intonatives;  $Z$ : score de la transformation de Fisher ;  $rms$  ( $rms > 0$ ): différence moyenne quadratique entre deux courbes intonatives. 102

Figure 25 : Comparaison de la courbe intonative du *mépris* avec celle de la *déclaration* et de la *politesse*. Les valeurs de F0 des contours sont représentées en demi-ton et normalisés à 10 points par voyelle.  $r$  ( $-1 < r < 1$ ): coefficient de corrélation pondéré entre deux courbes intonatives;  $Z$ : score de la transformation de Fisher ;  $rms$  ( $rms > 0$ ): différence moyenne quadratique entre deux courbes intonatives. 103

Figure 26 : Dendrogrammes montrant les distances entre les affects sociaux exprimés avec une phrase de neuf syllabes basées sur les coefficients de corrélation pondérés (en haut) et sur la différence moyenne quadratique du registre de F0 (en bas) entre les contours intonatifs des affects sociaux. 104

Figure 27 : Comparaison de la courbe intonative l'*ironie* avec celle de la *déclaration*, de l'*évidence*, de l'exclamation de *surprise neutre* et de *surprise positive*. Les valeurs de F0 des contours sont représentées en demi-ton et normalisés à 10 points par voyelle.  $r$  ( $-1 < r < 1$ ): coefficient de corrélation pondéré entre deux courbes intonatives;  $Z$ : score de la transformation de Fisher ;  $rms$  ( $rms > 0$ ): différence moyenne quadratique entre deux courbes intonatives. 105

Figure 28 : Comparaison de la courbe intonative de l'*irritation* avec celle de la *déclaration*, de l'*évidence*, de l'exclamation de *surprise neutre* et de la *question*. Les valeurs de F0 des contours sont représentées en demi-ton et normalisés à 10 points par voyelle.  $r$  ( $-1 < r < 1$ ): coefficient de corrélation pondéré entre deux courbes intonatives;  $Z$ : score de la transformation de Fisher ;  $rms$  ( $rms > 0$ ): différence moyenne quadratique entre deux courbes intonatives. 107

Figure 29 : Contours intonatifs de la *déception* et de la *résignation* comparés avec celui de la *déclaration*, ainsi que le contour intonatif du *doute* comparé avec celui de l'exclamation de *surprise neutre* et de la *question*. Les valeurs de F0 des contours sont représentées en demi-ton et normalisés à 10 points par voyelle.  $r$  ( $-1 < r < 1$ ): coefficient de corrélation pondéré entre deux courbes intonatives;  $Z$ : score de la transformation de Fisher ;  $rms$  ( $rms > 0$ ): différence moyenne quadratique entre deux courbes intonatives. 108

Figure 30 : Comparaison du contour intonatif de la *confiance* avec celui de la *déclaration* pour chaque longueur, et la comparaison des contours intonatifs entre la *confiance*, l'*évidence* et la *politesse* pour la phrase de 9 syllabes. Les valeurs de F0 des contours sont représentées en demi-ton et normalisés à

10 points par voyelle.  $r$  ( $-1 < r < 1$ ): coefficient de corrélation pondéré entre deux courbes intonatives;  $Z$ : score de la transformation de Fisher;  $rms$  ( $rms > 0$ ): différence moyenne quadratique entre deux courbes intonatives. 111

Figure 31 : Comparaison du contour intonatif de l'*autorité* avec celui de l'*irritation* pour chaque longueur. Les valeurs de F0 des contours sont représentées en demi-ton et normalisés à 10 points par voyelle.  $r$  ( $-1 < r < 1$ ): coefficient de corrélation pondéré entre deux courbes intonatives;  $Z$ : score de la transformation de Fisher;  $rms$  ( $rms > 0$ ): différence moyenne quadratique entre deux courbes intonatives. 112

Figure 32 : Comparaison du contour intonatif de la *déception* avec celui de la *résignation* pour chaque longueur. Les valeurs de F0 des contours sont représentées en demi-ton et normalisés à 10 points par voyelle.  $r$  ( $-1 < r < 1$ ): coefficient de corrélation pondéré entre deux courbes intonatives;  $Z$ : score de la transformation de Fisher;  $rms$  ( $rms > 0$ ): différence moyenne quadratique entre deux courbes intonatives. 112

Figure 33 : Comparaison du contour intonatif de l'*ironie* avec celui du *mépris* pour chaque longueur. Les valeurs de F0 des contours sont représentées en demi-ton et normalisés à 10 points par voyelle.  $r$  ( $-1 < r < 1$ ): coefficient de corrélation pondéré entre deux courbes intonatives;  $Z$ : score de la transformation de Fisher;  $rms$  ( $rms > 0$ ): différence moyenne quadratique entre deux courbes intonatives. 113

Figure 34: Taux de reconnaissance des 19 affects sociaux chinois pour les sujets chinois et français. La ligne pointillée rouge représente le seuil du hasard (5,3%). 118

Figure 35: Taux d'attractivité des 19 affects sociaux chinois pour les sujets chinois et français. La ligne pleine rouge représente le taux moyen d'attractivité calculé avec les 18 affects sociaux sauf la *déclaration*. 118

Figure 36: La différence de taux d'identification des 19 affects sociaux chinois entre les sujets chinois et français. 119

Figure 37: Effet de la modalité sur les taux d'identification des 11 affects sociaux chinois pour 32 sujets chinois. Les affects sociaux sont présentés en abscisse et les scores de reconnaissance en ordonnée. La ligne rouge pointillée indique le seuil du hasard (9%). 124

Figure 38: Effet de l'ordre de présentation des modalités sur les taux d'identification des 11 affects sociaux chinois pour 32 sujets chinois. Les affects sociaux sont présentés en abscisse et les scores de reconnaissance en ordonnée. La ligne rouge pointillée indique le seuil du hasard (9%). 125

Figure 39: Confusions des 11 affects sociaux chinois pour 32 sujets natifs dans la modalité « audio seul ». Les pourcentages d'identification sont indiqués dans les ovales et les taux de confusions sur les flèches. Seules les confusions qui affichent des scores supérieurs à deux fois le seuil du hasard sont rapportées. 127

- Figure 40: Classification hiérarchique des 11 affects sociaux chinois pour 32 sujets natifs dans la modalité « audio seul ». 128
- Figure 41: Confusions des 11 affects sociaux chinois pour 32 sujets natifs dans la modalité « vidéo seul ». Les pourcentages d'identification sont indiqués dans les ovales et les taux de confusions sur les flèches. Seules les confusions qui affichent des scores supérieurs à deux fois le seuil du hasard sont rapportées. 129
- Figure 42: Classification hiérarchique des 11 affects sociaux chinois pour 32 sujets natifs dans la modalité « vidéo seul ». 130
- Figure 43: Confusions des 11 affects sociaux chinois pour 32 sujets natifs dans la modalité « audio-vidéo ». Les pourcentages d'identification sont indiqués dans les ovales et les taux de confusions sur les flèches. Seules les confusions qui affichent des scores supérieurs à deux fois le seuil du hasard sont rapportées. 131
- Figure 44: Classification hiérarchique des 11 affects sociaux chinois pour 32 sujets natifs dans la modalité « audio-vidéo ». 132
- Figure 45: Taux d'identification des 19 affects sociaux chinois pour 30 sujets français et vietnamiens. Les affects sociaux sont présentés en abscisse et les scores de reconnaissance en ordonnée. La ligne rouge pointillée indique le seuil du hasard (5,3%). 144
- Figure 46: Taux d'identification des 19 affects sociaux chinois par 15 sujets français (FR) et 15 sujets vietnamiens (VN). Les affects sociaux sont présentés en abscisse et les scores de reconnaissance en ordonnée. La ligne rouge pointillée indique le seuil du hasard (5,3%). 145
- Figure 47: Taux d'identification des 19 affects sociaux chinois pour chaque énoncé par 15 sujets français (FR) et 15 sujets vietnamiens (VN). Les énoncés, présentés en abscisse, sont représentés par leur séquence tonale (1 représente le ton plat, 2 le ton montant, 3 le ton descendant-remontant, 4 le ton descendant et 0 le ton neutre); les scores de reconnaissance sont présentés en ordonnée. La ligne rouge pointillée indique le seuil du hasard (5,3%). 146
- Figure 48: Confusions perçues des 19 affects sociaux chinois pour 15 sujets français : les pourcentages d'identification sont indiqués dans les ovales et les taux de confusions sur les flèches. Seules les confusions qui affichent des scores supérieurs à deux fois le seuil du hasard sont rapportées. 149
- Figure 49: Confusions perçues des 19 affects sociaux chinois pour 15 sujets vietnamiens: les pourcentages d'identification sont indiqués dans les ovales et les taux de confusions sur les flèches. Seules les confusions qui affichent des scores supérieurs à deux fois le seuil du hasard sont rapportées. 150
- Figure 50: L'attractivité des 19 affects sociaux chinois – pourcentages cumulatifs de confusions provenant d'autres affects sociaux vers un affect social donné, pour les sujets chinois (CN), français (FR) et vietnamiens (VN). 151
- Figure 51: Classification hiérarchique des 19 affects sociaux chinois pour 15 sujets français (haut) et 15 sujets vietnamiens (bas). 152

- Figure 52: Distribution des 19 concepts d'affects sociaux par 120 sujets regroupés par langue maternelle dans les trois premières dimensions de l'AC. Les cercles autour de chaque concept indiquent un niveau de confiance de 97,5%. 160
- Figure 53: Distribution des 19 concepts d'affects sociaux par 120 sujets regroupés par genre dans les trois premières dimensions de l'AC. Les cercles autour de chaque concept indiquent un niveau de confiance de 97,5%. 161
- Figure 54: Classification hiérarchique des 19 affects sociaux à partir de leurs proximités conceptuelles pour les 120 sujets. 163
- Figure 55: Classifications hiérarchiques par langue maternelle des 19 concepts d'affects sociaux à partir de leurs proximités conceptuelles : sujets chinois (haut) et sujets français (bas). 164
- Figure 56: Contribution relative aux connaissances partagées par 120 sujets pour les 19 concepts d'affects sociaux due à ce qui est universellement partagé (Global), ce qui est spécifiquement lié à la langue et culture des sujets (Language), au genre des sujets (Gender) et à l'erreur d'échantillonnage (Chance). 165
- Figure 57: Distribution des 19 affects sociaux chinois présentés en concept et en audio sur les deux premières dimensions de l'AC. Les étiquettes en bleu correspondent aux concepts d'affects sociaux et les étiquettes en rouge aux stimuli auditifs; les cercles autour de chaque affect social indiquent un niveau de confiance de 97,5%. 169
- Figure 58: Distribution par genre des 19 affects sociaux chinois présentés en concept et en audio sur les deux premières dimensions de l'AC. Les étiquettes en bleu et en rouge correspondent respectivement aux stimuli auditifs d'affects sociaux perçus par les sujets masculins et par les sujets féminins ; les étiquettes en vert clair et en jaune correspondent respectivement aux concepts d'affects sociaux perçus par les sujets masculins et par les sujets féminins ; les cercles autour de chaque affect social indiquent un niveau de confiance de 97,5%. 170
- Figure 59: Dendrogramme de la classification hiérarchique des 19 affects sociaux chinois à partir de leurs proximités conceptuelles et psycho-acoustiques pour les 92 sujets natifs. 172
- Figure 60: Dendrogramme de la classification hiérarchique des 19 affects sociaux chinois à partir de leurs proximités conceptuelles et psycho-acoustiques pour les 92 sujets natifs, en fonction du genre. AA représente les stimuli auditifs, TT les concepts, H les sujets masculins et F les sujets féminins. 173
- Figure 61: Contribution relative aux connaissances partagées par 92 sujets natifs pour les 19 affects sociaux chinois apportée par ce qui universellement partagé (Global), ce qui est lié à la modalité de représentation des affects sociaux (Modality), au genre des sujets (Gender) et à l'erreur d'échantillonnage (Chance). 174
- Figure 62 : Effet de la modalité sur les taux d'identification des 11 affects sociaux français perçus par des sujets français. Les affects sociaux sont présentés en abscisse et les scores de reconnaissance en ordonnée. La ligne rouge pointillée indique le seuil du hasard (9%). 186



- Figure 63 : Effet de l'ordre de présentation des modalités sur les taux d'identification des 11 affects sociaux français perçus par des sujets français. Les affects sociaux sont présentés en abscisse et les scores de reconnaissance en ordonnée. La ligne rouge pointillée indique le seuil du hasard (9%). 187
- Figure 64 : Confusions des 11 affects sociaux français pour les 14 sujets français dans la modalité « audio seul ». Les pourcentages d'identification sont indiqués dans les ovales et les taux de confusions sur les flèches. Seules les confusions qui affichent des scores supérieurs à deux fois le seuil du hasard sont rapportées. 188
- Figure 65 : Classification hiérarchique des 11 affects sociaux français pour les 14 sujets français dans la modalité « audio seul ». 189
- Figure 66 : Confusions des 11 affects sociaux français pour les 14 sujets français dans la modalité « vidéo seul ». Les pourcentages d'identification sont indiqués dans les ovales et les taux de confusions sur les flèches. Seules les confusions qui affichent des scores supérieurs à deux fois le seuil du hasard sont rapportées. 190
- Figure 67 : Classification hiérarchique des 11 affects sociaux français pour les 14 sujets français dans la modalité « vidéo seul ». 191
- Figure 68 : Confusions des 11 affects sociaux français pour les 14 sujets français dans la modalité « audio-vidéo ». Les pourcentages d'identification sont indiqués dans les ovales et les taux de confusions sur les flèches. Seules les confusions qui affichent des scores supérieurs à deux fois le seuil du hasard sont rapportées. 192
- Figure 69 : Classification hiérarchique des 11 affects sociaux français pour les 14 sujets français dans la modalité « audio-vidéo ». 193
- Figure 70 : Effet de la modalité sur les taux d'identification des 11 affects sociaux français pour les 28 apprenants chinois du français. Les affects sociaux sont présentés en abscisse et les scores de reconnaissance en ordonnée. La ligne rouge pointillée indique le seuil du hasard (9%). 198
- Figure 71 : Effet de l'interaction entre l'ordre de présentation des modalités et le niveau de français sur les taux d'identification des 11 affects sociaux français pour les 28 apprenants chinois. Les affects sociaux sont présentés en abscisse et les scores de reconnaissance en ordonnée. La ligne rouge pointillée indique le seuil du hasard (9%). 199
- Figure 72 : Confusions des 11 affects sociaux français pour les 28 apprenants chinois du français dans la modalité « audio seul ». Les pourcentages d'identification sont indiqués dans les ovales et les taux de confusions sur les flèches. Seules les confusions qui affichent des scores supérieurs à deux fois le seuil du hasard sont rapportées. 200
- Figure 73 : Classification hiérarchique des 11 affects sociaux français pour les 28 apprenants chinois du français dans la modalité « audio seul ». 201

- Figure 74 : Confusions des 11 affects sociaux français pour les 28 apprenants chinois du français dans la modalité « vidéo seul ». Les pourcentages d'identification sont indiqués dans les ovales et les taux de confusions sur les flèches. Seuls les taux supérieurs à deux fois le seuil du hasard sont rapportés. 202
- Figure 75 : Classification hiérarchique des 11 affects sociaux français pour les 28 apprenants chinois du français dans la modalité « vidéo seul ». 203
- Figure 76 : Confusions des 11 affects sociaux français pour les 28 apprenants chinois du français dans la modalité « audio-vidéo ». Les pourcentages d'identification sont indiqués dans les ovales et les taux de confusions sur les flèches. Seules les confusions qui affichent des scores supérieurs à deux fois le seuil du hasard sont rapportées. 204
- Figure 77 : Classification hiérarchique des 11 affects sociaux français pour les 28 apprenants chinois du français dans la modalité « audio vidéo ». 205
- Figure 78 : Taux de reconnaissance de 11 affects sociaux français pour les sujets natifs et les apprenants chinois dans les trois modalités : « audio seul » (haut) ; « vidéo seul » (milieu) ; « audio-vidéo » (bas). 206
- Figure 79 : Effet de la modalité sur les taux d'identification des 11 affects sociaux chinois pour les 28 apprenants français du chinois mandarin. Les affects sociaux sont présentés en abscisse et les scores de reconnaissance en ordonnée. La ligne rouge pointillée indique le seuil du hasard (9%). 213
- Figure 80 : Confusions des 11 affects sociaux chinois pour les 28 apprenants français du chinois mandarin dans la modalité « audio seul ». Les pourcentages d'identification sont indiqués dans les ovales et les taux de confusions sur les flèches. Seules les confusions qui affichent des scores supérieurs à deux fois le seuil du hasard sont rapportées. 215
- Figure 81 : Classification hiérarchique des 11 affects sociaux chinois pour les 28 apprenants français du chinois mandarin dans la modalité « audio seul ». 216
- Figure 82 : Confusions des 11 affects sociaux chinois pour les 28 apprenants français du chinois mandarin dans la modalité « vidéo seul ». Les pourcentages d'identification sont indiqués dans les ovales et les taux de confusions sur les flèches. Seules les confusions qui affichent des scores supérieurs à deux fois le seuil du hasard sont rapportées. 217
- Figure 83 : Classification hiérarchique des 11 affects sociaux chinois pour les 28 apprenants français du chinois mandarin dans la modalité « vidéo seul ». 218
- Figure 84 : Confusions des 11 affects sociaux chinois pour les 28 apprenants français du chinois mandarin dans la modalité « audio-vidéo ». Les pourcentages d'identification sont indiqués dans les ovales et les taux de confusions sur les flèches. Seules les confusions qui affichent des scores supérieurs à deux fois le seuil du hasard sont rapportées. 219
- Figure 85 : Classification hiérarchique des 11 affects sociaux chinois pour les 28 apprenants français du chinois mandarin dans la modalité « audio-vidéo ». 220

Figure 86 : Taux de reconnaissance des 11 affects sociaux chinois pour les sujets natifs et les apprenants français dans les trois modalités différentes : « audio seul » (haut) ; « vidéo seul » (milieu) ; « audio-vidéo » (bas).

221

# Liste des tableaux

Table 1 : La distinction entre l'intonation expressive et l'intonation attitudinale, et quelques exemples d'étiquettes qui pourraient s'appliquer dans chaque cas (Wichmann, 2000).	20
Table 2 : Description des quatre tons chinois.	45
Table 3 : Affects sociaux en chinois et en français, ainsi que leur abréviation utilisée dans les travaux de thèse.	62
Table 4: Exemple du corpus des affects sociaux chinois, basé sur 152 énoncés des variations de la longueur, du placement des tons et de la structure syntaxique.	65
Table 5 : Sous-corpus utilisé dans l'expérience perceptive intra-culturelle pour les 19 affects sociaux chinois auprès de sujets natifs.	68
Table 6 : Effets des facteurs <i>Affect social</i> , <i>Genre</i> , <i>Longueur</i> et de leurs interactions sur la reconnaissance des affects sociaux chinois par des auditeurs natifs par un test ANOVA (effets significatifs en gras).	70
Table 7: Énoncés sélectionnés pour l'analyse acoustique des 19 affects sociaux chinois.	80
Table 8 : Indices prosodiques calculés sur la base des paramètres acoustiques de fréquence fondamentale, d'intensité et de durée.	82
Table 9 : Résultat du test de Bartlett et de Kaiser-Meyer-Olkin montrant l'adéquation des données à une ACP.	83
Table 10 : Proportion de la variance expliquée par chaque composante de l'ACP sur les 10 indices prosodiques.	84
Table 11 : Matrice des scores factoriels des 10 indices prosodiques après la rotation varimax.	84
Table 12 : Matrices de confusion des 19 affects sociaux à la longueur de 1, 2, 4 et 9 syllabes. Le taux de reconnaissance des affects sociaux est en gris foncé et les taux de confusion supérieurs à quatre fois le hasard (20%) sont en gris clair.	110
Table 13: Résultat de l'analyse de régression logistique mixte pas à pas descendante : les effets présentés dans le tableau sont testés comme <b>non significatifs</b> selon un test de rapport de vraisemblance ( $p > 0,05$ ).	123
Table 14: Sous-corpus utilisé dans l'expérience perceptive interculturelle de 19 affects sociaux chinois auprès de sujets français et vietnamiens.	142
Table 15: Effets des facteurs <i>Affect social</i> , <i>Langue</i> et <i>Ton</i> sur la reconnaissance des affects sociaux chinois par des auditeurs français et vietnamiens par un test ANOVA (effets significatifs en gras).	143
Table 16: Résultat des ANOVAs séparées par langue maternelle des sujets pour la reconnaissance des affects sociaux chinois (effets significatifs en gras).	144
Table 17: Taux d'identification des 19 affects sociaux pour les énoncés dont la dernière syllabe porte le ton 1, ton 2, ton 3, ton 4 ou le ton neutre, pour tous les sujets français et vietnamiens. Le taux moyen d'identification pour chaque énoncé est rapporté sur la dernière ligne.	147

Table 18: Bilan de classifications des affects sociaux chinois obtenues pour les sujets chinois, français et vietnamiens.	154
Table 19: Liste des étiquettes des 19 affects sociaux et de leur abréviation.	157
Table 20: Résultat de l'analyse de régression logistique mixte pas à pas descendante : les effets présentés dans le tableau sont testés comme significatifs selon un test de rapport de vraisemblance ( $p < 0,05$ ).	185
Table 21 : Résultat de l'analyse de régression logistique mixte pas à pas descendante : les effets présentés dans le tableau sont testés comme <i>non significatifs</i> selon un test de rapport de vraisemblance ( $p > 0,05$ ).	197
Table 22 : Résultat de l'analyse de régression logistique mixte pas à pas descendante : les effets présentés dans le tableau sont testés <i>non significatifs</i> au test de rapport de vraisemblance ( $p > 0,05$ ).	212

# Introduction

## Contexte de recherche

En tant qu'être social, chaque individu appartient à une communauté imprégnée d'une langue et d'une culture. Les individus interagissent dynamiquement entre eux par le langage, d'un côté pour créer et entretenir des relations intersubjectives, de l'autre côté parce qu'ils ont besoin d'échanger et de partager leurs connaissances, autrement dit, d'instruire et de s'instruire. La prosodie est le premier vecteur de langage que les humains utilisent d'abord en perception, et puis un peu plus tard en production, dès les premiers mois de leur vie (Dore, 1975; Halliday, 1975; Crystal, 1986; Konopczynski, 1990, 2000). Elle constitue bel et bien un facteur essentiel dans le développement des relations interpersonnelles. Elle renforce son rôle dans la communication orale avec ses fonctions plurielles selon une échelle de «linguisticité» : linguistiques, paralinguistiques et extralinguistiques. Outre sa participation à la construction de la chaîne parlée et du sens en contexte, la prosodie apporte également des informations sur l'état psychologique et affectif du locuteur (émotions, attitudes), ainsi que des indices sociolinguistiques (provenance géographique, âge, genre, classe sociale). La répartition de ces informations entre les interactants perdure lors de la communication parlée dont la réussite nécessite en revanche une bonne perception ou bien un décodage pertinent de ces messages transmis en para-verbal.

Or, par rapport aux informations sociolinguistiques, les expressions de l'affect du locuteur sont primordiales, et constituent une donnée permanente de la communication interpersonnelle, dans laquelle les actes de discours sont constamment médiatisés par la transmission et la perception de l'affect des participants. Les recherches sur les affects couvrent un champ très vaste et les variations constitutives de l'affect ne relèvent pas de la même nature dans la communication parlée.

En tant qu'héritière de Fónagy, Aubergé (2002a) s'est interrogée sur l'existence de différents contrôles de locuteur au processus cognitif de l'expression affective : les expressions involontairement contrôlées que sont les émotions et les expressions volontairement contrôlées que sont les affects sociaux. Ce point de vue correspond au

fond à la théorie de l'effet de *push* et l'effet de *pull* de (Scherer, 1984, 2001). Elle propose également d'élargir la notion d'attitude de Fónagy avec ses usages sociaux et concrétisés dans la communication interactionnelle. D'après elle, à part les attitudes strictement définies au sens linguistique et dans les actes de langage (par exemple le *doute* et l'*évidence*), il existe aussi des attitudes qui véhiculent des valeurs sociales hiérarchiques et définissent les rôles sociétaux des locuteurs pendant l'interaction (par exemple la *politesse* et l'*autorité*), et celles qui décrivent un contexte social général où se déroule l'interaction (par exemple la *séduction*, le *maternel* et l'*intimité*). Elle englobe les variations d'attitudes avec un terme plus général qui est « affects sociaux ». Les affects sociaux signalent l'intention du locuteur dans l'interaction. Cette intention est à priori définie par l'opinion, la croyance, l'affect et le comportement de l'individu et est également sous l'influence de son rôle sociétal et la situation communicative où il se trouve. D'un autre côté, les affects sociaux sont étroitement liés au langage et s'insèrent dans une culture donnée. Ils varient considérablement d'une culture à l'autre, même d'un individu à l'autre. Cela évoquerait éventuellement des difficultés ou des malentendus dans une communication interculturelle où les valeurs et les formes prosodiques de ces affects sociaux ne se partagent pas entre les locuteurs (Diaferia, 2002; Mac et al., 2010; Shochi et al., 2010; Aubergé, 2012; Rilliard et al., 2014). Par conséquent, l'apprentissage de la prosodie des affects sociaux d'une langue est évidemment nécessaire pour les apprenants en langue étrangère (LE) ou en langue seconde (L2) (Shochi et al., 2010).

Cependant, les manuels pédagogiques des langues étrangères proposent en général un travail sur l'identification et la production des phonèmes en langue cible et l'aspect strictement prosodique de la langue est souvent traité de façon superficielle, sans parler de la prosodie des affects sociaux. D'ailleurs, sous l'influence de sa langue maternelle, les apprenants LE/L2 s'approprient, au cours de leur apprentissage de la langue cible, une interlangue (un système de règles linguistiques abstraites propre aux apprenants LE/L2, qui constitue un stade intermédiaire entre la langue source et la langue cible), et transfèrent inconsciemment les caractéristiques phonético-phonologiques et prosodiques de leur langue maternelle sur la langue cible. Ceci les empêche de produire naturellement et de façon plus appropriée la langue étrangère. En l'occurrence, parallèlement à la correction phonétique qui est toujours prise en considération par les enseignants et a donné lieu à nombre de méthodes, la correction de la prosodie devrait

avoir également sa place dans l'enseignement/apprentissage des langues étrangères et secondes.

De plus, pour former une vraie compétence de communication en langue étrangère, il est loin d'être suffisant d'apprendre à construire des formes phonologiquement, lexicalement, morpho-syntaxiquement correctes dans la langue cible. Il n'y a pas de doute qu'il faut aussi savoir les co-construire avec des réalisations pertinentes de ces formes selon la nature et le contexte socioculturel de l'interaction et savoir exprimer, par la prosodie en particulier, son opinion, son intention, son attitude pendant l'acte de parole. Or, en comparaison avec l'enseignement du lexique, de la grammaire et de la syntaxe, celui de la prosodie ne pèse pas autant dans une classe de langue étrangère, et celui de la prosodie des affects sociaux est même très peu présent. Poursuivant la piste ouverte par Delattre (1966) et Fónagy (1983), nous posons notre regard sur ce champ peu exploité en espérant contribuer par nos connaissances et réflexions sur la science de parole et de prosodie au développement de l'enseignement de la prosodie socio-affective en langues étrangères.

### **Hypothèses et problématiques**

Ces travaux de thèse sont basés sur trois hypothèses principales :

Premièrement, en supposant que la prosodie fait partie de la construction du langage et que les valeurs des affects sociaux varient d'une langue/culture à une autre (Martins-Baltar, 1977; Fónagy, 1983; Fujisaki & Hirose, 1993; Wichmann, 2000), nous pensons que la forme prosodique des affects sociaux pourrait, pour une valeur analogue entre deux langues/cultures, être spécifique ou universelle. Par ailleurs, les expressions des valeurs d'attitude sont généralement acquises dans l'enfance (Clément, 1999) et elles pourraient aussi être apprises par des apprenants étrangers. Or, en considérant la situation de la langue chinoise et française, qui sont deux langues éloignées linguistiquement et culturellement, nous souhaitons mettre en lumière dans ces travaux de thèse les ressemblances et les contrastes de ces deux univers prosodiques dans leur expression des affects sociaux, en particulier sur quels points les « faux-amis » biaisent l'apprentissage de la prosodie socio-affective du français et du chinois LE et L2. De là, afin d'atteindre ces objectifs, nous avons mené une série d'expériences perceptives à la croisée des deux langues.



En dépit de cet intérêt porté à l'enseignement de la prosodie des affects sociaux en LE ou L2, ces travaux de thèse se positionnent effectivement sur le champ de la recherche sur la parole et la prosodie au lieu d'aller plus avant dans le domaine didactique ou pédagogique. Ils se donnent comme objectif de fournir des données de première main à la recherche sur la méthode d'enseignement/apprentissage de la prosodie des affects sociaux du français pour les apprenants chinois et du chinois pour les apprenants français.

La deuxième hypothèse porte sur la représentation cognitivo-culturelle des affects sociaux. Chaque langue/culture attribue des étiquettes vernaculaires aux concepts attitudinaux, Schubiger (1958) et O'Connor & Arnold (1978) ont utilisé près de 300 étiquettes pour nommer les attitudes au sens psychologique. A l'instar de Fónagy (1983), des étiquettes naïves des attitudes françaises ont été utilisées dans l'enseignement de la langue. Ce sont des étiquettes empiriques mais convenables à l'usage pragmatique et vernaculaire. En enrichissant et en étendant la liste d'étiquettes de Fónagy avec des étiquettes d'affects sociaux, nous nous permettons de nous interroger sur deux questions qu'il n'a pas abordées : les affects sociaux sont-ils conceptuellement universels (dans notre étude, nous nous intéressons uniquement aux langues chinoise et française) ? Dans une langue et culture, la catégorisation conceptuelle des affects sociaux converge-t-elle vers la catégorisation perceptive ?

En partant de ces questions, au lieu d'entamer les problèmes catégoriel et dimensionnel des affects sociaux, nous envisageons de mesurer et comparer la distance cognitive des étiquettes d'affects sociaux dans la langue chinoise et française, et pour aller plus loin, d'examiner la divergence entre la distance cognitive des affects sociaux et leur distance acoustique s'il en existe.

L'autre hypothèse est formulée sur le rôle du genre dans la fonction des affects sociaux. Quand on parcourt l'histoire humaine chronologiquement ou synchroniquement, on constate que l'homme et la femme n'ont pas le même rôle dans les sociétés ni la même représentation aux yeux de l'un et de l'autre, et qu'ils utilisent des stratégies différentes pour exprimer leur émotion, intention ou attitude (Loveday, 1981; Ohara, 1992). En ce qui concerne les émotions, quelques différences, discutables et discutées, ont pu être mises en lumière dans quelques expériences (Audibert et al., 2008). Ce qui nous amène à

vérifier s'il existe également des différences entre les genres en terme de perception des affects sociaux.

Au départ de ces hypothèses, nous nous focalisons dans ces travaux de thèse sur les affects sociaux, qui sont volontairement contrôlés par le locuteur durant son processus cognitif pour mettre en valeur son intention, son point de vue, son rôle sociétal et la relation sociale avec son interlocuteur. Ces travaux sont réalisés sur la base d'un corpus contrôlé d'affects sociaux en chinois mandarin et en français. L'objectif de ces études porte sur quatre aspects des affects sociaux prototypiques : le comportement perceptif des natifs et des non natifs, les principales caractéristiques prosodiques, la dispersion cognitive et psycho-acoustique des affects sociaux, et la perception multimodale des affects sociaux chinois et français auprès d'apprenants en LE ou L2.

### **Démarche générale et organisation de la thèse**

Après une présentation brève de l'état de l'art, nous orienterons notre recherche sur une étude contrastive de la prosodie des affects sociaux du français et du chinois, qui constituera une première étape pour des recherches spécifiques orientées vers la conception et l'application de méthodes d'enseignement de la prosodie socio-affective en français et en chinois langue étrangère ou seconde.

Dans le premier chapitre, nous passerons en revue les grandes théories de l'émotion et de l'attitude en psychologie sociale, et nous introduirons ensuite une vision linguistique et phonostylistique sur l'attitude, plus largement l'affect social, qui est l'objet de notre recherche. Par la suite, nous présenterons les fonctions essentielles de la prosodie dans le système de communication parlée et les premières réflexions sur la prosodie des attitudes dans l'enseignement de langues. Finalement, nous exposerons quelques études sur la prosodie des expressions émotionnelles et attitudinales en chinois.

Ensuite, nous présenterons les travaux empiriques sur des affects sociaux du chinois et du français que nous avons menés dans le cadre méthodologique réalisé et développé par Grépillat (1996) et Morlec (1997) pour 6 attitudes françaises, par Diaferia (2002) pour 11 attitudes anglaises, par Shochi (2008) pour 12 attitudes japonaises et puis par Mac (2012) pour 16 affects sociaux vietnamiens.

En premier lieu, nous aborderons la perception auditive et multimodale de 19 affects sociaux chinois, avec pour objectif de mettre en lumière le comportement perceptif des sujets natifs (chapitre 2). Nous avons commencé par construire un corpus chinois en considérant à la fois des combinaisons tonales jouant sur la longueur variable des énoncés et la modulation des formes morpho-syntaxiques. L'idée de concevoir ce corpus est venue de la tentative d'éliminer l'influence linguistique vis-à-vis de l'influence de la prosodie lors de l'énonciation. Les énoncés utilisés sont neutres, à savoir qu'ils ne contiennent aucun mot ou aucune expression qui implique une émotion ou une attitude. Par ailleurs, la focalisation n'a pas été prise en compte dans ce corpus. Par la suite, nous avons validé auditivement le corpus auprès de sujets natifs, et parallèlement, la contribution de deux modalités communicationnelles, l'audition et la vision, a aussi été mesurée dans une autre expérience multimodale. Nous présenterons également une analyse acoustique sur un corpus sélectionné qui avait été validé dans l'expérience perceptive, afin d'extraire les caractéristiques principales de la prosodie des affects sociaux chinois et d'essayer d'examiner les similarités entre les contours intonatifs des expressions socio-affectives.

En second lieu, nous regarderons la perception de la prosodie socio-affective du chinois, ainsi que le traitement cognitif des affects sociaux dans un cadre interculturel (chapitre 3). Nous avons d'abord effectué un test perceptif auprès de sujets français (d'une langue/culture lointaine du chinois) et de sujets vietnamiens (d'une langue à tons relativement proche du chinois) dans le but d'examiner l'effet potentiel du ton sur la perception de la prosodie des affects sociaux chinois. Puis, nous avons examiné le concept d'affect social en mesurant la distance conceptuelle entre les étiquettes chez les sujets chinois et français, ainsi que la distance psycho-acoustique des affects sociaux en chinois mandarin. L'objectif de cette étude conceptuelle consiste à répondre à deux questions posées comme hypothèse : est-ce que les affects sociaux se représentent cognitivement de la même façon pour des langues différentes ; est-ce qu'il existe un décalage entre le concept et la perception d'un affect social au sein d'une langue. En conséquence, deux expériences spécifiques ont été effectuées pour vérifier ces hypothèses.

Enfin, le dernier chapitre de cette thèse est consacré à une étude sur la perception multimodale des affects sociaux français et chinois respectivement par des apprenants

chinois et français (chapitre 4). Nous décrirons dans ce chapitre l'évaluation de la compétence à reconnaître les affects sociaux des apprenants avec différents degrés d'acquisition de la langue cible. En comparant les résultats des sujets étudiants étrangers avec ceux des sujets natifs, nous essayerons d'un côté, de déceler les affects sociaux problématiques qui poseront éventuellement des problèmes aux apprenants étrangers dans la communication face-à-face en langue cible, et de l'autre côté, de donner des suggestions envisageables pour l'enseignement de la prosodie des affects sociaux français et chinois langue étrangère ou seconde.

# **Chapitre 1. Analyse de l'état de l'art**

## **1.1. Introduction**

Ce premier chapitre a pour objectif de poser et de définir les concepts nécessaires à nos études intra- et interculturelles sur la perception et la conceptualisation des affects sociaux dans les chapitres 2 et 3, ainsi qu'à l'étude contrastive de la perception multimodale des affects sociaux du chinois et du français auprès d'apprenants en langue étrangère ou seconde dans le chapitre 4.

La première partie de ce chapitre est consacrée à une présentation des théories et des points de vue des affects, précisément de l'émotion et de l'attitude aussi bien en psychologie qu'en linguistique et phonostylistique. Puis, dans la deuxième partie, nous abordons la notion et les fonctions de la prosodie, son rôle dans la communication parlée, ainsi que des études antérieures sur l'enseignement de la prosodie attitudinale. La dernière partie propose une brève présentation du chinois mandarin et de sa prosodie, et présente une rapide synthèse des travaux antérieurs sur la prosodie émotionnelle et attitudinale en chinois mandarin.

## **1.2. Quelques notions sur les affects**

Avant que la recherche sur les affects humains devienne scientifique dans des domaines concernés (physiologie, biologie, psychologie, neuropsychologie ou linguistique), le terme de « passion » a été généralement utilisé pour désigner tous les affects. Au fur et à mesure du développement de la science affective, il a été constaté que, dans de nombreux ouvrages sur les affects, l'émotion et l'attitude sont confondues et utilisés comme synonymes. Ceci pourrait être dû à la difficulté de les définir et de les évaluer de manière objective. Cependant, il nous paraît nécessaire et utile de passer en revue les études effectuées sur l'émotion et l'attitude dans les domaines psychologique, linguistique, ainsi que phonostylistique, afin de les cerner et distinguer.

### **1.2.1. Affects en psychologie**

Nous allons d'abord survoler quelques grandes théories sur l'émotion et l'attitude en psychologie.

#### 1.2.1.1. Emotions en psychologie

Il faut remonter à l'antiquité pour trouver les origines de l'étude des émotions chez des grands maîtres de la philosophie, tels que Platon et Aristote. Mais c'est grâce aux psychologues que l'étude des émotions est finalement entrée dans le champ scientifique à la fin du 19<sup>e</sup> siècle. Cornelius (1996, 2000) a proposé une distinction entre quatre perspectives théoriques contemporaines en fonction de leur façon de définir, d'étudier et d'expliquer les émotions. Il s'agit des perspectives évolutionniste, physiologique, cognitive et socio-constructiviste.

##### 1.2.1.1.1. Perspective évolutionniste

C'est dans l'ouvrage de « L'expression des émotions chez les hommes et les animaux » rédigé en 1872 que Darwin a élaboré pour la première fois ses pensées évolutionnistes sur les émotions. Les émotions sont, pour lui, universelles et adaptatives. Selon lui, l'universalité des émotions se traduit par le fait qu'elles sont censées ne pas être limitées géographiquement et culturellement, et la nature adaptative des émotions permet à l'individu de répondre de façon appropriée aux problèmes causés par l'environnement extérieur.

A l'instar des travaux de Darwin, ses successeurs ont aussi cherché à démontrer l'universalité des expressions faciales des émotions en dépit de la culture (Ekman & Friesen, 1978; Ekman, 1989, 1999). Dans les années 1960, beaucoup de chercheurs ont essayé d'illustrer l'existence de certaines émotions considérées comme fondamentales et universelles pour l'individu, à savoir les émotions de base ou primaires. Il y a au moins cinq émotions fondamentales : la joie, la tristesse, le dégoût, la colère et la peur. Le nombre des émotions primaires varie et monte jusqu'à huit ou dix selon les auteurs. Mais les six émotions de base identifiées par Ekman (1989) (joie, tristesse, dégoût, colère, peur et surprise) sont les plus fréquemment citées dans la littérature, qui sont même appelées « Big Six » par Cornelius (2000). Les émotions complexes comme l'anxiété ou l'enthousiasme sont considérées par eux, comme provenant d'une combinaison des émotions de base. Ils ont aussi précisé que les émotions sont déclenchées par des stimuli provenant des événements extérieurs et remplissent une fonction adaptative en servant de signal pour l'organisme.

Mais au fil du temps, certains chercheurs en émotions ont également dépassé les pensées novatrices de Darwin. Ils ont ancré leurs études sur l'idée que les émotions remplissent une autre fonction régulatrice à part celle d'adaptation, pour que l'individu puisse non seulement s'adapter aux exigences environnementales mais aussi réguler ses comportements selon les critères sociaux. Cette perspective de l'émotion, représentée par Wallon (1934/1970), Malrieu (1952) et Dumas (1948), se centre plutôt sur les codes et les expressions émotionnelles des individus et sur les fonctions adaptatives des émotions. Puis un accord consensuel a été décidément pris sur la fonction régulatrice et adaptative de l'émotion chez la plupart des théoriciens contemporains.

#### 1.2.1.1.2. Perspective physiologique

Elle est aussi appelée « la théorie de l'émotion de James-Lange ». Par rapport à Darwin, William James s'attache plus à la nature de l'émotion (James, 1884, 1894) en considérant l'activation physiologique comme le déclencheur principal des processus émotionnels. En illustrant son point de vue par un exemple de peur, James a expliqué l'origine de l'émotion avec la perception des changements corporels aux niveaux musculaire et neurovégétatif, qui proviennent du système nerveux périphérique. L'importance a été donnée par l'auteur aux réactions périphériques lors du déclenchement d'une émotion qui est équivalente au sentiment subjectif. Une année après la publication des idées de James, Lange (1885/1967) a exprimé ses affinités avec la théorie de James et on appelle ainsi cette théorie, la théorie de l'émotion « James-Lange », considérée comme la première théorie scientifique de l'émotion.

Pourtant, cette théorie a été critiquée par Cannon (1927) et de Bard (1928) dans les années 20. D'après eux, c'est le système thalamique (système nerveux central) qui détermine le déclenchement d'une émotion spécifique, mais pas les réponses neurovégétatives. Cela signifie que, pour Cannon et Bard, l'évaluation cognitive est le vrai déclencheur central des émotions et les réactions physiologiques ne sont que la conséquence des émotions.

Bien que les approches de James-Lange et de Cannon-Bard poursuivent toutes la piste évolutionniste en approuvant la nature adaptative des émotions à l'environnement, les deux approches mettent en évidence une différence majeure avec l'approche darwinienne: leur intérêt porte plutôt sur les émotions elles-mêmes que sur leurs

expressions. Cependant, si l'on observe une grande divergence entre ces deux approches, on peut remarquer également une opposition radicale sur la question de la séquence temporelle qui sépare la perception des changements corporels de l'émotion et de la localisation des structures cérébrales impliquées dans l'émotion (Tcherkassof, 2009, cité par Nugier, 2009). Cette controverse a suscité des études sur l'importance de la « cognition » dans l'émotion.

#### 1.2.1.1.3. Perspective cognitive

Cette perspective est considérée comme la théorie dominante de toutes les perspectives sur les émotions. Les premières œuvres sur les théories de l'évaluation cognitive de l'émotion datent du début des années soixante du XX<sup>e</sup> siècle, et sont représentées par les travaux de Arnold (1960) et de Lazarus (1966). Arnold (1960) a suggéré pour la première fois que la signification d'un événement déclenchant une émotion est établie par un processus d'évaluation (« appraisal » en anglais) de l'événement. Lazarus (1966) a développé cette idée en proposant l'existence d'un autre processus de réévaluation (« reappraisal ») qui pourrait modifier les premières impressions ainsi que l'émotion déclenchée. Ils sont les premiers à utiliser le concept d'appraisal ou d'évaluation cognitive.

D'après les tenants de ces théories, les émotions que l'individu ressent dans une situation seraient déclenchées et différenciées par l'évaluation qu'il attribue à la situation, et en accompagnement d'une tendance à réagir sur les aspects spécifiques isolés et évalués par lui-même. C'est donc une évaluation subjective, directe, immédiate, non-réflexive et intuitive, d'un certain nombre de critères ou dimensions de la situation, qui serait pertinente pour l'individu (Grandjean & Scherer, 2009). Les différentes émotions peuvent ainsi être décrites en termes de patterns d'évaluations cognitives différentes.

En fait, depuis les années 80, beaucoup d'études ont été réalisées pour spécifier les dimensions d'évaluation et la relation entre l'évaluation et l'expérience d'émotions particulières. Bien qu'elles se différencient au niveau de l'identification et du nombre de dimensions, on pourrait trouver des similarités observables. Grandjean & Scherer (2009) ont remarqué que, d'une façon générale, les dimensions d'évaluation les plus consensuellement incluses dans les recherches sont au nombre de 4 : la détection de la



pertinence, l'évaluation de l'implication, le potentiel de maîtrise et l'évaluation de la signification normative. Actuellement, les théories de l'évaluation cognitive de l'émotion sont censées être une référence pour expliquer le mécanisme de genèse et la différenciation des émotions. Parallèlement, elles sont particulièrement utiles à expliquer les différences intra- et inter- individuelles, notamment culturelles et sociales dans le déclenchement et la différenciation des émotions.

#### 1.2.1.1.4. Perspective socio-constructiviste

A l'instar des théories de l'évaluation cognitive de l'émotion qui se penchent sur la subjectivité et les variables culturelles auxquelles sont liées les différences des émotions, les tenants de la perspective socio-constructiviste de l'émotion considèrent les émotions comme produits d'une culture donnée, qui pourraient être acquises par le biais de la socialisation de l'individu et renforcées au travers de son rôle social.

Contrairement aux théories Darwinienne et Jamesienne, cette perspective considère que la plupart des comportements, attitudes, états de l'être humain sont des constructions purement sociales et culturelles (Gergen, 1985), ainsi ils devraient varier en termes de codes socioculturels partagés dans les diverses communautés. De l'autre côté, elle s'identifie et se différencie de l'approche d'appraisal de l'émotion en remarquant que le même événement provoque probablement des émotions différentes chez les membres d'un groupe à l'autre, même s'il s'agit de la même culture, car ces groupes sociaux ne demeurent pas identiques en fonction de la classe sociale, de la génération ou de l'affiliation politique.

Les théoriciens socio-constructivistes de l'émotion, représentés par Averill (1980), ont expliqué les émotions comme des scripts applicables conformément aux normes socioculturelles de référence et qui apparaîtraient de façon transitoire selon l'exigence de situation. A leur avis, les émotions des individus proviennent plutôt de leur façon d'interpréter les liens qui unissent la situation, le système de valeurs et les référents culturels. Les émotions sont donc, en quelque sorte, dépendantes de la culture. C'est aussi pour cette raison que les socio-constructivistes insistent pour étudier les différences entre les émotions appartenant à des cultures différentes au niveau du contenu plutôt que de se concentrer seulement sur les similitudes ou sur les émotions de base comme le font les évolutionnistes des années 60.

Dans ce contexte, de nombreuses études ont été menées sur les étiquettes linguistiques des émotions dans des cultures différentes, car « les émotions peuvent prendre des étiquettes différentes selon les cultures, être regroupées sous une même étiquette alors qu'elles sont conceptuellement différentes, ou encore être étiquetées dans une langue et pas dans une autre » (Nugier, 2009). Ces études sur le langage constituent le principal support empirique de la perspective socio-constructiviste, et cette réflexion sur le lien entre l'étiquette sémantique, le concept et la culture nous a inspiré une étude spécifique sur les concepts des affects sociaux, que nous allons présenter dans le chapitre 3.

#### 1.2.1.2. Attitudes en psychologie sociale

Voyons dans cette partie la notion de l'attitude en psychologie sociale, ainsi que quelques modèles classiques de la structure d'attitude, que sont le modèle unidimensionnel, multidimensionnel et le modèle tripartite révisé.

##### 1.2.1.2.1. Notion de l'attitude

L'apparition de la notion d'attitude en psychologie sociale remonte aux travaux de Thomas et Znaniecki (1918-1920) qui ont initié de nouvelles recherches en psychologie sociale en y introduisant le concept d'attitude. A travers l'observation des paysans polonais immigrés aux Etats-Unis, ils ont indiqué que le comportement individuel n'est pas explicable en termes fonctionnels, et que pour le comprendre il faut savoir comment la personne définit la situation à laquelle elle peut être confrontée. Cette définition dépend également de l'attitude qu'elle a à l'égard des objets qui constituent l'environnement. Ils étaient convaincus que les attitudes sont des dispositions mentales explicatives du comportement et pourraient expliquer les réactions de l'individu devant les stimulations sociales.

Il s'en est suivi des façons différentes de définir l'attitude chez les psychologues sociaux et sociologues pendant plusieurs décennies et parmi les multiples définitions du concept d'attitude, celle proposée par Allport (1935) est la plus couramment acceptée. Selon lui, l'attitude est un état mental et neurologique, qui permet à l'individu de prévoir son comportement dans une situation donnée à l'égard d'un stimulus donné. Il s'agit donc de la préparation à l'action envers un objet ou un événement sur la base de l'expérience de

l'individu<sup>1</sup>. L'auteur a montré dans cette définition, d'un côté, son désir de conceptualiser les attitudes par leur contenu ou leur nature (état mental et neurologique), de l'autre côté, l'intention de distinguer l'attitude de la notion d'instinct avec laquelle elle pourrait être confondue. Dix ans après, Krech & Crutchfield (1948) ont défini l'attitude comme organisation durable du processus cognitif<sup>2</sup> en soulignant son lien étroit avec le comportement de l'individu. Quelques sociologues et psychologues ont même pensé que le comportement de l'individu pourrait varier selon la spécificité situationnelle (Fuson, 1942; Campbell, 1950).

Il existe quasiment autant de définitions du concept d'attitude que d'auteurs qui ont travaillé sur les attitudes. Certains auteurs accordent un consensus à la polarité positive ou négative et la définissent comme la seule dimension de l'attitude ; d'autres auteurs ne limitent pas leur définition à cette seule évaluation, ils affirment que l'attitude est une disposition résultant de l'organisation de trois composantes. Il s'agit du modèle multidimensionnel de l'attitude.

#### 1.2.1.2.2. Modèle unidimensionnel classique

C'est le modèle d'attitude le plus courant qui affirme que l'attitude représente la réponse évaluative d'un objet en terme d'attirance (favorable) ou de répulsion (défavorable). L'objet est ainsi placé par l'attitude sur la position du continuum d'évaluation, et la réponse évaluative est associée en fait à l'affect de l'individu. Cette propriété affective fait rapprocher de ce modèle unidimensionnel les approches cognitives selon lesquelles l'attitude est juste un affect associé à la représentation mentale d'un objet stockée en mémoire (Fazio et al., 2004; Nosek et al., 2005).

#### 1.2.1.2.3. Modèle multidimensionnel ou modèle tripartite

Le modèle tripartite (ou modèle à trois parties) de l'attitude est un modèle psychologique classique du concept d'attitude, qui préconise les trois composantes d'attitude. Proposé par Rosenberg & Hovland (1960), ce modèle considère l'attitude comme une disposition à trois composantes : cognitive, affective et comportementale

---

<sup>1</sup> "An attitude is a mental and neural state of readiness, organized through experience, exerting a directive and dynamic influence upon the individual's response to all objects and situations with which it is related" (Allport, 1935, p. 810).

<sup>2</sup> An attitude can be defined as an enduring organization of motivational, emotional, perceptual, and cognitive processes with respect to some aspect of the individual's world" (Krech & Crutchfield, 1948, p. 152).

(ou conative). La composante cognitive est liée aux croyances ou opinions évoquées par l'objet d'attitude ; la composante affective est associée à l'affect suscité ; et la composante conative renvoie au plan d'action (Vallerand & Lafrenaye, 2006). Bien que les trois dimensions soient nettement distinguées par les tenants de cette approche, elles devraient se corrélérer entre elles parce qu'elles reflètent la même attitude. Breckler (1984) a démontré les corrélations entre les trois composantes à travers des expérimentations avec différents objets attitudinaux. Par ailleurs, ses recherches ont également montré que chaque composante possédait une contribution unique au cours de la définition de l'attitude, et ceci suggérerait une validité discriminante qui permettrait de distinguer chacune d'elles.

Zanna & Rempel (1988) ont proposé d'enrichir le modèle de Rosenberg & Hovland en intégrant le modèle unidimensionnel. Selon ce modèle révisé, l'attitude est définie, en premier lieu, comme une catégorisation de l'objet attitudinal en fonction de la dimension évaluative bipolaire. L'attitude est donc un jugement favorable ou défavorable envers l'objet. En second lieu, l'attitude est considérée comme un élément résultant d'une convergence de trois sortes d'information : une information cognitive, une information affective et une information conative. Il y a par conséquent quatre composantes distinctes dans le modèle tripartite révisé : l'attitude, jugement sur l'objet ; l'affect, associé à l'émotion ressentie ; les croyances, qui sont les conséquences négatives ou positives associées à l'objet ; et la structure cognitive d'anticipation de l'action (Vallerand & Lafrenaye, 2006).

Sur la base de ce modèle, Zanna et Rempel ont suggéré que les trois sortes d'informations pouvaient déterminer l'attitude soit séparément soit conjointement. Cela signifie qu'un même objet pourrait évoquer des attitudes différentes selon le type d'information saillante dans des situations différentes. En utilisant ce modèle, Zanna et al. (1990) ont démontré que les déterminants fonctionnent différemment chez les individus : pour certaines personnes, l'attitude provient de certains déterminants et pour d'autres, plutôt d'autres déterminants.

En comparaison avec le modèle tripartite classique de Rosenberg & Hovland, qui considère les trois composantes comme conséquences de l'attitude, le modèle tripartite révisé montre qu'au lieu de résulter du construit hypothétique de l'attitude, les expériences cognitives, affectives et comportementales de l'individu déterminent plutôt

son attitude envers un objet. Ce modèle de structuration de l'attitude met en évidence les caractéristiques de l'attitude et sa relation avec l'émotion : la dimension affective relève de l'origine émotive de l'attitude, mais les deux autres dimensions l'en séparent et impliquent une prise de conscience.

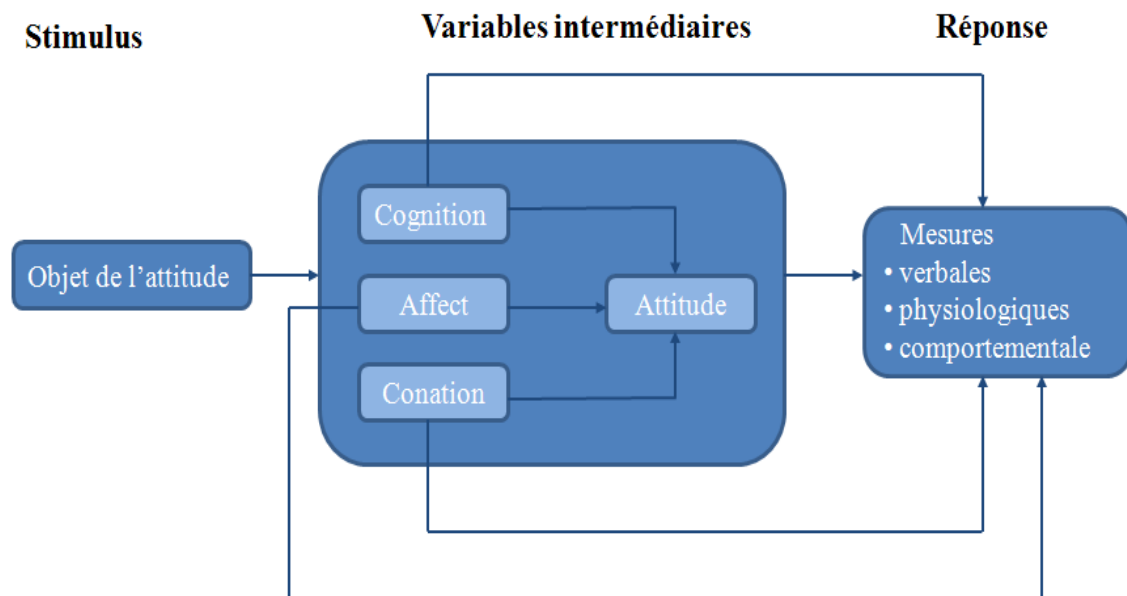


Figure 1: Représentation du modèle tripartite révisé de Zanna et Rempel (Vallerand & Lafrenaye, 2006, p. 244).

### 1.2.2. Affects en linguistique et en phonostylistique

Après avoir passé brièvement en revue les grandes théories de l'émotion et de l'attitude en psychologie, nous tentons ici de regarder les émotions et les attitudes d'un point de vue linguistique et phonostylistique. Néanmoins, il faut élucider préalablement la notion de l'attitude en linguistique, parce qu'elle n'est pas identique à celle en psychologie sociale.

Les psychologues qui se sont intéressés à la communication affective proposent de distinguer les émotions brutes et les émotions socialisées (*social emotion* en anglais) (Arndt & Janney, 1991). C'est ce que les linguistes désignent par l'émotion et l'attitude. Les émotions brutes, ou l'émotion pour les linguistes, se réfèrent à des réactions instinctives de l'organisme chez l'individu et sont des réponses involontaires aux stimuli externes. Par contre, les émotions socialisées, ou l'attitude pour les linguistes, sont des expressions contrôlées de l'individu à l'égard de son environnement, relevant de conventions partagées par les membres d'une communauté linguistique donnée (Di

Cristo, 2013, p. 236). Dans le cadre de notre recherche sur les attitudes du chinois et du français dans la communication parlée, nous nous référons plutôt à l'attitude en termes linguistiques et phonostylistiques qu'en termes psychologiques.

Voyons ici comment les approches linguistique et phonostylistique regardent les émotions et les attitudes, à savoir comment les distinguer et les catégoriser sur le plan linguistique.

#### 1.2.2.1. Distinction entre l'émotion et l'attitude

Du fait que l'expression des émotions et l'expression des attitudes se réalisent pareillement sur le plan physique (par exemple l'expression faciale et le geste) et langagier (langage et parole), elles se font objet de recherche chez des théoriciens de la communication parlée, notamment phonostylisticiens, phonéticiens et pragmaticiens.

Dans ses travaux de 1971 et 1983, Iván Fónagy a proposé le fameux principe de double encodage pour mettre en relief des aspects linguistiques (accentuation, rythme, intonation) et paralinguistiques (émotions et attitudes), qui sont impliqués par tout acte de parole (Fónagy, 1971, 1983). Selon cette théorie, le premier encodage correspond à la double articulation définie par Martinet (1970) comme des monèmes composés de phonèmes, tandis que le second encodage prend en charge la transformation (ou la distorsion) et la concrétisation du message linguistique produit dans le premier encodage, de même que la stratégie d'interaction sociale du sujet parlant. Or, aux yeux de l'auteur, l'expression des émotions primaires et l'expression des attitudes ne relèvent pas de la même nature, bien qu'elles fassent partie toutes du message secondaire, affectif et gradient. A l'encontre de l'expression émotionnelle qui est spontanée, l'expression attitudinale est préfabriquée, et plus conventionnée que l'expression d'émotions primaires (Fónagy, 1956). Il a également distingué les émotions et les attitudes sur le plan segmental et prosodique dans les langues en observant l'expression du doute et la question simple en russe (Fónagy, 1969): les émotions primaires se manifestent simultanément à tous les niveaux de l'appareil vocal, sous-glottique, glottique et supraglottique. Pourtant, les attitudes sont réalisées essentiellement au niveau glottique.

A l'exemple du double encodage de Fónagy, Ohala (1996) a distingué, d'un point de vue éthologique, les différents niveaux de construction du message qui transmet des

informations affectives. Dans ce modèle, en dehors du contenu du message proprement dit, l'émotion et l'attitude constituent deux niveaux différents :

*At a fundamental level people recognize signals that convey messages where the transmission of the message itself has survival value. [...] Such signals should be found cross culturally and, perhaps cross species. [...] Such signals would be motivated more by the effect they would have on receivers than by the inner (psycho)physiological state of the signaler.*

*At another level would be signals which do reflect the (psycho)physiological state of the signaler. In most cases, arguably, displaying these conditions does not serve the best interests of the signaler but they presumably arise from physiological states which are themselves beneficial, i.e., adaptive, responses to external conditions.*

*At another level one can recognize ways that humans convey their attitudes about the receiver, about the content or referent of their utterance, or about themselves. This would include irony, sarcasm, indifference, and the like. Such messages do not confer obvious survival benefit to the signaler and are probably acquired, i.e., learned. Thus they are likely to vary considerably from culture to culture and perhaps even from one individual to another.*

D'un point de vue phonostylistique, Léon (1993) situe l'émotion et l'attitude sur une échelle de stylisation. L'émotion brute, indice d'un désordre et d'un choc, est spontanée et fugitive. Quand l'émotion brute est bien contrôlée et que ses traits prosodiques deviennent plus importants que les traits paralinguistiques, elle est alors stylisée et on passe aux stades différents que sont le sentiment et l'attitude. Le sentiment est un stade relativement stable, durable et atténué par rapport à l'émotion brute, il est un signe plus contrôlé, consciemment ou non, donc plus motivé que l'émotion. L'attitude constitue un stade bien stable et est consciemment contrôlée par le locuteur. Elle est sémiotiquement considérée comme un signe conventionnel partagé dans une communauté linguistique donnée.

#### 1.2.2.2. Quelques points de vue sur la catégorisation des attitudes

Les premières tentatives de définir ou catégoriser les étiquettes d'attitude utilisées par les linguistes ont trouvé leur origine dans les réflexions sur le lien entre l'intonation et les affects que l'intonation pouvait transmettre. Martins-Baltar (1977) et Fónagy et al. (1983) ont déjà distingué dans leurs travaux les attitudes propositionnelles (les attitudes postulées sur des propositions) et les attitudes sociales (attitudes choisies en fonction du contexte et adressées à un interlocuteur), alors que Couper-Kuhlen (1986, pp. 185-187), a défini l'attitude comme le comportement du locuteur, opposée à l'émotion qui est son état psychologique. Cette définition d'attitude a exclu les attitudes cognitives des psychologues sociaux, qui sont chargées d'exprimer l'opinion, la croyance ou la connaissance du locuteur. Les dernières sont également nommées attitudes propositionnelles, différentes des attitudes de Couper-Kuhlen, à savoir les attitudes de comportement qui correspondent au fond aux attitudes sociales de Martins-Baltar (1977) et de Fónagy et al. (1983).

D'un point de vue pragmatique, Wichmann (2000) a expliqué la distinction des deux catégories d'attitudes avec leur relation avec l'interaction conversationnelle : les attitudes propositionnelles, comme les émotions, sont indépendantes de l'interaction, et elles peuvent se produire sans l'existence d'un interlocuteur ; les attitudes définies comme le comportement de locuteur sont pourtant dépendantes de l'interaction et ne peuvent pas être présentes hors d'une situation conversationnelle<sup>3</sup>. Différentes des attitudes propositionnelles et des émotions, les attitudes de comportement, aussi nommées attitude 'énonciative' par Martins-Baltar (1977), sont considérées comme une expression adaptative au contexte et à la situation de communication à la fois au niveau textuel et intonatif (Martins-Baltar, 1977, p. 70). A partir de cette distinction, Wichmann a proposé de classer comme intonation 'expressive' les caractéristiques intonatives qui transmettent les émotions pures et les attitudes propositionnelles, et comme intonation 'attitudinale' les indices linguistiques et les indices non-linguistiques de la voix, qui

---

<sup>3</sup> "In terms of conversational interaction this means that they (propositional attitudes) can exist without another person being present. Attitude defined as speaker behavior, on the other hand, cannot be present independently of interaction" (Wichmann, 2000).



reflètent le comportement du locuteur dans une situation donnée, soit comme prévu par le locuteur lui-même, soit comme déduit par l'auditeur, ou les deux<sup>4</sup>.

Expressive intonation reflets...	and...	Attitudinal intonation reflets...
	propositional attitude	speaker behavior
emotion	Opinion, belief, knowledge about a person or issue	intended and/or perceived in a given context
He is (feeling).  <i>happy</i> <i>angry</i> <i>sad...</i>	I am  <i>critical</i> <i>impressed</i> <i>disapproving...</i>	You are being  <i>condescending</i> <i>friendly</i> <i>rude...</i>

Table 1 : La distinction entre l'intonation expressive et l'intonation attitudinale, et quelques exemples d'étiquettes qui pourraient s'appliquer dans chaque cas (Wichmann, 2000).

Sur la base de cette distinction catégorielle des attitudes, de Moraes et al. (2010) ont proposé dans leur étude sur les attitudes du portugais brésilien de classer 13 étiquettes d'attitude classées en deux catégories en fonction de la modalité d'énonciation que sont l'assertion et l'interrogation. Toutes les attitudes sociales (au nombre de 6) sont compatibles avec les deux modalités, et il n'y a que deux sur sept des attitudes propositionnelles qui s'adaptent aux deux modalités. La plupart d'attitudes propositionnelles dépendent plutôt de la modalité de la phrase, et c'est le cas du doute, de l'évidence et de l'ironie pour l'assertion et de la confirmation et de la rhétoricité pour l'interrogation.

- *Modalité assertive :*
  - *Attitudes sociales : arrogance, autorité, mépris, irritation, politesse et séduction ;*
  - *Attitudes propositionnelles : incrédulité, surprise, doute, évidence et ironie ;*
- *Modalité interrogative :*
  - *Attitudes sociales : arrogance, autorité, mépris, irritation, politesse et séduction ;*
  - *Attitudes propositionnelles : incrédulité, surprise, confirmation et rhétoricité.*

<sup>4</sup> "I will categorise as 'expressive' intonation those intonational characteristics which appear to convey pure emotion, and emotions arising from or closely linked to beliefs, knowledge and opinion. I use 'attitudinal' intonation to mean any intonational cue which, together with non-linguistic information in the voice, reflects speaker behavior in a given situation, either as intended by the speaker, or as inferred by the receiver, or both" (Wichmann, 2000).

### 1.2.2.3. Des attitudes aux affects sociaux

Après avoir fait un survol de quelques grandes théories de l'émotion et de l'attitude en psychologie sociale et linguistique, nous voudrions présenter ici une nouvelle vision phonostylistique sur le concept d'attitude, proposée par Aubergé (2002a).

Inspirée par les idées de Fónagy (1983), Aubergé (2002a) s'est d'abord interrogée sur l'existence de différents contrôles de locuteur au processus cognitif de l'expression affective. Elle a proposé de distinguer les expressions émotionnelles et les expressions attitudinales selon la nature du contrôle (involontaire vs. volontaire) exercé par le locuteur pendant la communication (voir aussi Léon, 1993). Selon elle, les émotions sont déclenchées par un contrôle involontaire et sont exprimées à travers des canaux divers qui pourraient être aussi bien la voix que le visage. Concernant la construction ontogénétique, les émotions sont innées chez l'individu et se présentent avec peu de variantes dans les différentes cultures humaines. A l'opposé des émotions, les attitudes émergent plutôt de manière volontaire et intentionnelle. Elles sont déclenchées par une prise de contrôle volontaire sur le contrôle involontaire des émotions, à savoir par une simulation d'une émotion, mais participant en même temps à l'organisation de l'énonciation sur laquelle porte l'émotion. Cette distinction entre contrôle involontaire et volontaire correspond à celle de l'effet de *push* et de *pull* proposée par (Scherer, 1984, 2001).

Parallèlement, en reprenant la notion d'attitude de Fónagy (1983), l'auteur a également proposé de l'élargir en tenant compte des valeurs sociales que les attitudes prennent dans la communication interactionnelle. D'après elle, à part les attitudes strictement définies au sens linguistique et dans les actes de langage (par exemple le *doute* et l'*évidence*), il existe aussi des affects qui véhiculent des valeurs sociales hiérarchiques et des informations sur le rôle sociétal des locuteurs pendant l'interaction (par exemple la *politesse* et l'*autorité*), et ceux qui décrivent un contexte social à long termes qui permet aux sujets leur nature plus endogène de leur relation (par exemple la *séduction*, le *maternel* et l'*intimité*). Elle a ainsi proposé d'englober ces expressions affectives avec un terme plus général qui est « affects sociaux ». Les affects sociaux mettraient en circulation entre les interactants les informations sur la situation d'énonciation et leur relation sociale, et aideraient les interlocuteurs à effectuer leur acte illocutoire et perlocutoire dans les tours de parole. Il est à signaler que pour l'auteur, les deux

modalités d'énonciation de base (déclaration et interrogation) font aussi partie des attitudes (Aubergé et al., 1997). Le fait que le sujet n'exprime aucune attitude met juste en œuvre une attitude qui consiste à montrer que le sujet n'a pas une attitude particulière, ou qu'il ne veut pas ou ne peut pas l'exprimer dans son discours (Danes, 1994; Aubergé et al., 1997; Morlec et al., 2001; Aubergé, 2002a; Shochi et al., 2005).

Etant donné leur lien étroit avec la société et la culture, les affects sociaux varient considérablement d'une culture à l'autre, même d'un individu à l'autre, et montrent souvent des spécificités en concept comme en réalisation prosodique en fonction de la langue et de la culture, mais cela n'exclut pas quelques universalités des affects sociaux dans toutes les langues. Par contre, plus d'attention devrait être attribué aux spécificités des affects sociaux dans les cultures différentes, car les spécificités culturelles évoquerait éventuellement des difficultés ou des malentendus dans la communication interculturelle où les valeurs et les formes prosodiques de ces affects sociaux ne se partagent pas entre les locuteurs (Shochi et al., 2009; Mac et al., 2010). En conséquence, la prosodie des affects sociaux devrait être apprise non seulement par les enfants en tant que partie intégrante de leur langue maternelle (Clément, 1999; Roseberry-Mckibbin & Brice, 1999), mais aussi par les apprenants en langue étrangère et seconde (Abelin & Allwood, 2000), pour améliorer leur compétence de communication en langue cible.

Pour conclure, nous nous intéressons dans cette thèse aux affects sociaux, qui sont des expressions affectives conventionnellement encodées dans une culture et une langue, et qui transmettent des informations sur la situation communicative et la relation sociale interpersonnelle. Ils pourraient être appris par des apprenants étrangers si les valeurs d'affects sociaux ne se partagent pas entre la langue cible et la langue source. Dans le cadre des travaux de thèse, au lieu d'attaquer les problèmes catégoriels et dimensionnels des affects sociaux, nous nous concentrons sur la perception des affects sociaux dans un contexte interculturel et didactique afin de proposer des suggestions à l'apprentissage de la prosodie socio-affective en français et en chinois langue étrangère ou seconde.

#### 1.2.2.4. Etiquettes données aux émotions et aux affects sociaux

A l'époque où l'émotion et l'attitude n'étaient pas bien différenciées, les étiquettes données aux émotions et aux attitudes variaient souvent selon les auteurs. Or, au fur et à

mesure de l'évolution de la recherche en communication expressive, l'émotion et l'attitude sont définies de plus en plus précisément et reçoivent chacune un système d'étiquettes différent. Comme la *joie*, la *tristesse*, la *colère*, la *peur* et le *dégoût* sont communément utilisés comme des termes émotionnels, il existe aussi des étiquettes beaucoup utilisées pour désigner les attitudes : près de 300 étiquettes ont été utilisées par Schubiger (1958) et O'Connor & Arnold (1978) pour nommer les attitudes au sens psychologique, par exemple: *abrupt*, *accusant*, *affable*, *affecté*, *affectueux*, *agressif*, *agréable*, *amusé*, *colère*, *hostile*, *agacé*, *antagoniste*, *apologétique*, *émouvant*, *appréciatif*, *appréhensif*, *approbatif*, *ergoteur*, *arrogant*, *impressionné*, *autoritaire*...<sup>5</sup>.

Sur la base des distinctions entre les émotions et les attitudes, ou plus largement les affects sociaux, que nous venons de décrire ci-dessus, nous voulons enrichir ici la liste d'attitudes de Fónagy en en ajoutant quelques unes. Pour nous, le *doute*, la *politesse*, le *mépris*, la *surprise*, l'*évidence*, l'*autorité*, etc. sont bel et bien des étiquettes évaluées dans le cadre des affects sociaux.

Un autre problème plus critique concernant les étiquettes des émotions et des affects sociaux porte sur la relation entre les termes sémantiques et la conceptualisation du monde.

Introduite par l'hypothèse de Sapir et Whorf dans les années 50, la réflexion sur la langue et la conceptualisation du monde était au cœur de l'intérêt des linguistes et des anthropologues, qui se sont séparés dans deux camps : l'un a prétendu que les termes sémantiques de couleurs sont sujets aux contraintes sémantique universelles, représenté par Berlin & Kay (1969), Bornstein et al. (1976) et Kay & Maffi (1999) ; l'autre, représenté par Saunders (1995) Lucy (1997) et Wierzbicka (2006), a insisté sur l'idée que les contraintes linguistiques et culturelles existent toujours.

Pour éviter le problème causé par la traduction des étiquettes émotionnelles d'une langue à l'autre, Wierzbicka (1996) a proposé d'utiliser une métalangue sémantique naturelle (MSN). Par exemple, pour décrire le concept de *joie*, au lieu d'utiliser les mots tels que « *joy* » en anglais, ou « *heureux* » en français, ou bien « *glücklich* » en allemand, on pourrait utiliser une formule sémantique universelle :

---

<sup>5</sup> En anglais : 'abrupt, accusing, affable, affected, affectionate, aggressive, agreeable, airy, amused, angry, animated, annoyed, antagonistic, apologetic, appealing, appreciative, apprehensive, approving, argumentative, arrogant, awed, authoritative...'

*X feels something*

*Sometimes people think something like this:*

*something very good happened to me*

*I wanted this*

*everything is good*

*I can't want anything else*

*because of this, they feel something good*

*X feels like this*

Sans aucun doute le paradigme MSN a réussi à mettre en relation l'universalité et les spécificités culturelles de chaque langue en termes d'émotion, et il s'applique même à divers domaines scientifiques en langue : sémantique lexicale, linguistique interculturelle, sociolinguistique, etc. (Durst, 2003, p. 181). Rilliard et al. (2014) ont repris ce paradigme dans leur étude de la perception interculturelle des expressions de politesse et d'impolitesse en japonais pour définir les concepts des attitudes en question.

Or, dans notre travail sur l'enseignement des affects sociaux prosodiques en langues étrangères, au lieu d'utiliser des scripts comme MSN, nous avons recours à une autre façon de définir les concepts d'affects sociaux. Il s'agit de la description d'un contexte et d'une situation communicative précise où un certain affect social pourrait avoir lieu. Tous les concepts sont aussi représentés par des étiquettes vernaculaires dans une langue donnée. Plus de détails seront donnés dans le chapitre 4.

### **1.3. Survol des théories sur la prosodie**

L'étude de la prosodie connaît une longue histoire, Di Cristo (1975) a déjà référencié dans son ouvrage « Soixante et dix ans de recherches en prosodie » plus de 4000 entrées bibliographiques relatives à la prosodie. Aujourd'hui, l'essor des travaux sur la prosodie témoigne encore de son dynamisme et de ses enjeux pour des domaines divers : linguistique, analyse de discours, communication affective, apprentissage des langues, technologie vocale, recherche clinique, etc. Dans cette section, nous allons d'abord présenter la notion générale de prosodie ainsi que ses paramètres acoustiques. Ensuite,

nous entamerons les fonctions plurielles de la prosodie dans les différentes perspectives et son rôle essentiel dans la communication parlée, et aborderons ensuite son importance dans l'apprentissage des langues, surtout dans le contexte de l'enseignement du français en Chine.

### 1.3.1. Notion générale de la prosodie

Sur le champ de la langue et de la parole, l'emploi du terme « prosodie » varie souvent selon les auteurs d'une extrémité où la prosodie ne traite que des phénomènes en grammaire des langues anciennes (latin et grec), tels que la quantité syllabique et l'accent tonique, à l'autre extrémité où ce terme est utilisé à révéler les réalisations sonores qui « n'entre(nt) pas dans le cadre de la 'segmentation' monématique et phonématique » du langage oral (Martinet, 1969). Le terme de prosodie est donc réservé comme synonyme des traits suprasegmentaux des énoncés.

Néanmoins, il est difficile de définir la prosodie dans l'absolu. Du point de vue linguistique, Crystal (1969) l'a définie comme un signe des distinctions communicatives au niveau lexical et grammatical, ainsi que des attributs émotionnels de la parole; Vaissière (2006) l'a considérée comme l'ensemble des phénomènes d'accentuation lexicale et d'intonation, et comme les facteurs de performance. D'un point de vue acoustique, Crary & Tallman (1993) ont déclaré que la prosodie correspond plutôt aux caractéristiques suprasegmentales de la parole, qui concernent la fréquence fondamentale, la durée et l'intensité, lesquelles fonctionnent ensemble à former la mélodie de la parole (Crary & Tallman, 1993, p. 245). D'un autre point de vue concernant ses fonctions, la prosodie peut être définie par la multiplicité de ces trois paramètres qui opèrent à différents niveaux: lexical, démarcatif, pragmatique (intentions et attitudes), comportemental, émotionnel et idiolectal (âge, sexe et identité du locuteur).

Vu la complexité de la prosodie qui recouvre divers phénomènes langagiers, aucune définition de la prosodie serait universellement valable pour tous les chercheurs sans précision de la théorie ou de l'approche dans laquelle s'intègre la prosodie. Dans notre travail, nous voudrions retenir la définition synthétique de la prosodie donnée par (Di Cristo, 2013):

*La prosodie est une branche de la linguistique consacrée à l'analyse des propriétés formelles (phonologie prosodique), de la matérialité (phonétique prosodique) et de la fonctionnalité des éléments non verbaux de l'expression orale, non coextensifs aux phonèmes, tels que l'accent, les tons, l'intonation, la quantité, le tempo et les pauses, que l'on qualifie collectivement de prosodèmes. La matérialité physique des prosodèmes se manifeste par les variations de la fréquence fondamentale (F0), de la durée et de l'intensité (paramètres prosodiques acoustiques), ces variations étant perçues par l'auditeur comme des changements respectifs de hauteur, de longueur et de volume sonore (paramètres prosodiques auditifs). Les informations prosodiques ont la particularité d'être polysémiques et de véhiculer à la fois des informations d'ordre linguistique, paralinguistique et extralinguistique, qui se révèlent déterminantes pour l'interprétation des énoncés du discours et de la conversation. (Di Cristo, 2013, p. 21).*

### 1.3.2. Paramètres de la prosodie

Bien qu'il y ait des façons différentes de considérer la prosodie et qu'il soit par conséquent impossible de se référer à une définition strictement consensuelle, on s'est mis d'accord sur les trois paramètres acoustiques de la prosodie sur le plan phonétique que sont la fréquence fondamentale, la durée et l'intensité:

- La fréquence fondamentale (F0), définie comme la fréquence de vibration périodique des cordes vocales, correspond à la fréquence laryngée mesurée à partir de la décomposition harmonique du signal selon le théorème de Fourier. L'unité de mesure de la fréquence fondamentale est le Hertz (Hz), qui signifie le nombre de périodes de vibration des cordes vocales par seconde (Lacheret-Dujour & Beaugendre, 1999; Martin, 2009).
- La durée syllabique est considérée comme l'intervalle de temps nécessaire pour émettre le signal de parole. Précisément, elle est l'organisation temporelle du message, qui concerne le débit de parole (nombre de syllabes réalisées par seconde), le tempo (accélération ou ralentissement du débit dans un groupe prosodique) et les pauses (Lacheret-Dujour & Beaugendre, 1999, p. 12). Il est

plus difficile de mesurer la durée phonémique que la durée syllabique. La durée se mesure en seconde(s) ou en millièmes de secondes (ms).

- L'intensité a trait à l'énergie ou l'effort vocal accordé dans le signal sonore, elle résulte de l'amplitude des cordes vocales, de la force du flux d'air expiré et des variations de fréquence (Lacheret-Dujour & Beaugendre, 1999). C'est une mesure logarithmique de l'énergie du signal, dont l'unité est le décibel (dB). Le seuil différentiel d'intensité pour la parole est d'environ 3 dB, autrement dit, il faut doubler la puissance si l'on veut augmenter une voyelle de 3 dB (Rossi, 1999, p. 206). Les mesures d'intensité n'ont de sens qu'en valeurs relatives, par exemple entre deux voyelles consécutives (Martin, 2009, p. 17)

Bien que ces trois paramètres soient étroitement liés l'un à l'autre en raison des contraintes physiologiques, la fréquence fondamentale se présente, dans la majorité des recherches sur la prosodie, comme un paramètre essentiel qui constitue par excellence le support physique de l'intonation. Elle demeure également, par rapport aux autres paramètres prosodiques, celui qui est le plus étudié dans la communication parlée des états psychologiques (Lieberman & Michaels, 1962; Halliday, 1970; O'Connor & Arnold, 1978), parce qu'elle est l'indice le plus significatif et le plus important dans la perception de la prosodie (Cruttenden, 1986). Différente de la fréquence fondamentale et de l'intensité dont le changement est modulé par les activités phonatoires, la durée dépend plutôt des activités articulatoires et constitue le support physique temporel du signal de la parole. Elle est de même une composante essentielle de la prosodie, dont l'importance est cruciale (Di Cristo, 2013, p. 89). L'intensité est un paramètre qui, contrairement à la fréquence fondamentale et à la durée, ne fait pas l'objet de soins attentifs dans la littérature prosodique à cause de l'éventuelle difficulté à la mesurer et à l'évaluer objectivement (Di Cristo, 2013, p. 89). La fréquence fondamentale et la durée sont plus faciles à mesurer et à percevoir, donc plus significatives pour la recherche sur la prosodie (Benesty, 2008).

Au niveau linguistique, l'intonation, le rythme, le ton et l'accent proviennent de la modulation des trois premiers paramètres physiques. Certains théoriciens incluent le ton et l'accent dans la prosodie, alors que d'autres (Aubergé et al., 1997) considère qu'ils relèvent de l'ordre lexical. Il faut signaler que notre présent travail de recherche sur la prosodie des affects sociaux se positionne dans cette dernière perspective.



A part les trois paramètres acoustiques conventionnels de la prosodie, la qualité de voix constitue sans doute un aspect important de la prosodie. Elle a notamment été rattachée à l'expression vocale des affects, en plus de véhiculer aussi des informations extralinguistiques telles que l'âge et le sexe du locuteur. La qualité vocale est vue comme le quatrième paramètre de la prosodie par Campbell & Mokhtari (2003)<sup>6</sup>. Elle est considérée comme porteuse remarquable d'informations sur l'état émotionnel du locuteur (Scherer, 1986; Johnstone & Schere, 1999; Gobl & Ní Chasaide, 2000; Gobl & Ní Chasaide, 2003). Abercrombie (1967) et Laver (1980) ont donné des descriptions de la qualité de voix comme suit :

*The term "voice quality" refers to those characteristics which are present more or less all the time that a person is talking: it is a quasi permanent quality running through all the sound that issues from the mouth (Abercrombie, 1967, p. 91).<sup>7</sup>*

*Voice quality is conceived here in a broad sense, as the characteristic auditory colouring of an individual speaker's voice, and not in the more narrow sense of the quality deriving solely from laryngeal activity. Both laryngeal and supralaryngeal features will be seen as contributing to voice quality (Laver, 1980, p. 1)<sup>8</sup>.*

Il est évident que dans le sens général la qualité de voix se charge de « colorer » la voix de l'individu qui diffère de celle des autres. Elle varie souvent avec les états affectifs du locuteur, à savoir ses états émotionnels ou attitudeaux (Aubergé, 2002b; Shochi et al., 2006). Elle se fait donc un paramètre fondamental pour exprimer les affects du locuteur (Gobl & Ní Chasaide, 2003), mais pas le seul, car les trois autres dimensions de la prosodie interviennent aussi et fonctionnent ensemble avec la qualité de la voix (Bänziger & Scherer, 2003; Audibert et al., 2006; Mozziconacci, 2001). En associant les

---

<sup>6</sup> Campbell & Mokhtari (2003) ont montré que le *Normalized Amplitude Quotient* (NAQ), mesure mise au point par Alku et al. (2002), varie notamment en fonction de l'interlocuteur et du type d'acte de parole. Ce NAQ constitue une dimension des paramètres de la source glottique, et est donc l'un des composants complexes de la qualité vocale. Cette mesure est corrélée à la perception d'une voix plus ou moins tendue et donc à la dimension tendue/relâchée de la source glottique (voir d'Alessandro, 2006).

<sup>7</sup> Le terme de « qualité de voix » fait référence aux caractéristiques à plus ou moins tout le temps durant lequel une personne parle : c'est une qualité quasi permanente dans tous les bruits sortis de sa bouche. (Traduit par l'auteur).

<sup>8</sup> La qualité de voix est conçue ici dans un sens large, comme la caractéristique auditive colorée par la voix d'un locuteur, mais pas dans un sens plus étroit de la qualité qui dérive uniquement de l'activité du larynx. Les traits laryngés et supralaryngés contribuent tous les deux à la qualité de voix. (Traduit par l'auteur).

mouvements des muscles du larynx (les différents types phonatoires) aux qualités de voix résultantes, Laver (1980) a suggéré qu'en anglais la voix soufflée (« *breathy voice* ») est liée à l'intimité, la voix chuchotée (« *whispery voice* ») avec la confidentialité, la voix rauque (« *harsh voice* ») avec la colère et la voix craquée (« *creaky voice* ») avec la résignation causée par l'ennui.

### 1.3.3. Fonctions plurielles de la prosodie

Très tôt dans l'histoire de la psycholinguistique, des constatations remarquables sur le lien entre l'état cognitif et la parole ont pu montrer que non seulement la prosodie joue un rôle fonctionnel dans l'architecture de la parole, mais également sa structure a une fonction fondamentale dans la reconnaissance de la parole. Sous l'angle linguistique, la prosodie assume des fonctions syntaxiques : les informations sur la modalité syntaxique de l'énoncé sont transmises en premier lieu par le profil prosodique global de l'énoncé ; en second lieu, l'intonation permet d'indiquer, à l'intérieur de l'énoncé, la hiérarchisation des constituants de la construction morpho-syntaxique de l'énoncé. Sous l'angle expressif, la prosodie se charge de l'expression de l'état affectif du locuteur, qui est souvent une combinaison de l'état émotionnel et attitudinal (Campbell, 2005). Sous l'angle cognitif, la prosodie fournit des indices à l'auditeur pour qu'il puisse segmenter le continuum sonore et identifier les unités linguistiques, et parallèlement, reconnaître les informations importantes du message.

#### 1.3.3.1. Du point de vue de la linguisticité

Crystal (1975) a proposé de classer les fonctions de la prosodie selon une échelle de linguisticité. Selon cette échelle, les fonctions les plus linguistiques sont représentées par les fonctions d'organisation et de mise en forme du matériau verbal (mots, phrases, textes, etc.) ; les fonctions les moins linguistiques, par les fonctions d'identification des individus et de leur phonostyle. A partir de cette conception, nous allons présenter dans cette partie les fonctions de la prosodie en suivant la piste qui, à l'instar de Léon (1971) est consensuellement acceptée : la distinction des fonctions linguistique, paralinguistique et extralinguistique de la prosodie.

#### Les fonctions linguistiques

Les fonctions linguistiques (ou l'intonation) sont les fonctions les plus couramment étudiées parmi toutes les fonctions de la prosodie. D'un point de vue classique, la

prosodie est considérée comme assumant deux fonctions linguistiques : une fonction de segmentation/hiérarchisation (également appelée fonction de démarcation ou fonction de structuration de l'énonciation), avec laquelle la prosodie sert à segmenter le continuum sonore de l'énoncé en plusieurs sections qui se structurent hiérarchiquement (Daneš, 1960, p. 43). Dans le discours, les signes prosodiques prennent en charge la segmentation des composants syntaxiques : par exemple, la hauteur (la fréquence fondamentale) du discours est généralement haute au début et descendante à la fin de la phrase (Grosz & Hirschberg, 1992; Caelen-Haumont & Bel, 2000). En réalité, cette fonction structurale, qui est observable dans toutes les langues, permet à la prosodie de réaliser l'organisation du matériau verbal et se place ainsi au sommet de la hiérarchie des fonctions linguistiques assumées par la prosodie (Di Cristo, 2013, p. 176).

Outre les fonctions de segmentation et de hiérarchisation, la prosodie assume également une fonction de focalisation (ou fonction accentuelle, fonction de saillance) qui prend en charge la mise en valeur de certaines parties de l'énoncé (Vaissière, 1997, p. 54).

En termes de fonctions linguistiques, la prosodie remplit également une fonction distinctive lexicale (Di Cristo, 2013). Cette fonction lexicale varie sur deux plans :

- l'accentuation lexicale pour les langues à l'accent libre (par exemple, l'anglais, l'italien et l'espagnol), qui prend en charge l'identification des morphèmes constitutifs du mot et met en relief les morphèmes accentogènes ;
- le ton lexical pour les langues à tons (par exemple, le chinois mandarin, le vietnamien et le thaï), dont la mise en œuvre contribue à « accroître la capacité du lexique d'une langue et à actualiser des distinctions morphologiques » (Di Cristo, 2013, p. 176). Le nombre des tons lexicaux varie selon la langue ; dans une langue à tons donnée, ils possèdent chacun un contour prosodique spécifique, ce qui leur permet de distinguer la signification lexicale d'un mot qui porte toujours un ton.

### Les fonctions paralinguistiques

Outre les fonctions linguistiques que la prosodie remplit en participant à la structuration de la langue, elle joue parallèlement un rôle paralinguistique (ou ectolingustique).

Les fonctions paralinguistiques contribuent à véhiculer des informations sur l'état émotionnel ou l'attitude du locuteur. Le codage des émotions et des attitudes en

linguistique est appelé la fonction expressive du langage, ou bien « fonction d'expression de l'affect de la prosodie » selon Di Cristo. Par ailleurs, les processus prosodiques de l'expression affective se caractérisent par des aspects motivés et conventionnels. L'expression prosodique d'un affect donné (par exemple « colère ») réfère au signe arbitraire (le mot « colère ») et en même temps fait partie de l'affect lui-même (la colère), ou de l'action que le signe dénote (Fónagy, 1983, p. 149). D'un autre côté, cette expression est également liée à des conventions linguistiques et s'avère par conséquent conventionnelle dans la mesure où elle peut se soustraire à certaines normes d'usage (Lacheret-Dujour & Beaugendre, 1999, p. 61).

En combinant la prosodie émotionnelle et la prosodie attitudinale avec le terme général de prosodie affective, Campbell (2005) a ainsi indiqué : « *When coupled with gestures, affective prosody imparts vitality to discourse and greatly influences the content and impact of the message* »<sup>9</sup>. De même Ross (2000) a proposé le même sujet : « *The paralinguistic features of language, as exemplified by affective prosody, may thus play an even more important role in human communication than the exact choice of words* »<sup>10</sup>.

### Les fonctions extralinguistiques

L'autre rôle joué par la prosodie dans la communication interindividuelle est un rôle extralinguistique (ou idiolectale).

Différentes des fonctions paralinguistiques, les fonctions extralinguistiques fournissent des informations sur le profil du locuteur : âge, sexe, origine régionale, appartenance socio-éducative, catégorie socio-professionnelle, etc., parmi lesquelles le sexe nous apparaît un facteur prépondérant qui influence le comportement langagier de l'individu.

Durant l'évolution de la société humaine, le rôle des hommes et des femmes n'est pas toujours identique et varie d'une culture à l'autre. Ainsi certaines langues/cultures, comme le japonais, ont des systèmes prosodiques spécifiques pour les hommes vs. les femmes (Sadanobu, 2004). L'identité des deux sexes est en effet dans une large mesure liée aux stéréotypes selon lesquels on est élevé et cultivé ainsi que jugé par les autres. Par exemple, le genre masculin pourrait nous permettre de penser à la grandeur, à la

---

<sup>9</sup>. Associée avec les gestes, la prosodie affective donne la vitalité au discours et influence fortement le contenu et l'impact du message. (Traduit par l'auteur)

<sup>10</sup> Exemplifiés par la prosodie affective, les traits paralinguistiques de la langue peuvent donc jouer un rôle encore plus important dans la communication humaine que le choix exact de mots. (Traduit par l'auteur)

force, à la violence, à la conscience, à la pensée, à la volonté, tandis que le genre féminin à l'inverse évoquerait la passivité, la dépendance et l'absence d'individualité, auxquelles sont associés la petitesse, la beauté, la grâce, le charme, la coquetterie, la séduction dangereuse, la douleur, la tendresse, la peur, la versatilité, la déloyauté, etc. (voir Michard, 2002, p. 47-48).

Certes, beaucoup de ces stéréotypes ont changé avec le temps dans certaines cultures comme la culture chinoise et la culture française. Les deux pôles de sexes se rapprochent de plus en plus, toutefois, nous sommes d'avis que les changements sociaux ne suffisent pas à effacer radicalement les images stéréotypées représentant les deux sexes, les hommes et les femmes ne s'identifient toujours pas de manière identique sur le plan socioculturel et sur le plan linguistique.

Des études précédentes ont déjà montré que le sexe, qui participe à la personnalité de l'individu, constitue un facteur essentiel qui influence la façon dont l'échange langagier se déroulera ou sera compris. Par exemple : une attitude comme la séduction, qui est bien reconnue en perception croisée japonais/anglais/anglais américain/français, est perçue correctement mais jamais produite en contexte par un homme japonais (Shochi et al., 2007), alors qu'elle est produite et perçue par les femmes japonaises. Mac et al. (2010) ont montré que la parole adressée à un jeune enfant est très mal identifiée quand le locuteur est un homme, alors que des pré-tests semblent montrer que des formes analogues produites par une femme sont perçues correctement. Loveday (1981) et Ohara (1992) ont également indiqué que les hommes et les femmes japonais utilisent des stratégies différentes pour exprimer leur émotion, intention ou attitude. Au niveau de la phonologie, Trudgill (1983) a indiqué que plus de femmes prononcent le « r » après la voyelle que les hommes, quand il s'agit de l'un des signes distinctifs des personnes bien cultivées et occupant une place socialement prestigieuse.

Inspirés par les recherches effectuées sur la corrélation entre le genre et la langue, notamment entre le genre et l'attitude prosodique, nous avons l'intention de vérifier dans notre travail s'il existe de même des différences entre homme et femme en termes de perception des affects sociaux chinois.

#### 1.3.3.2. Du point de vue fonctionnel

Aubergé (2003) a proposé de distinguer dans son modèle prosodique trois fonctions différentes de la prosodie : les fonctions linguistiques, les fonctions des émotions et les fonctions des attitudes.

En ce qui concerne les fonctions linguistiques, elle a indiqué que en dehors des fonctions de segmentation / hiérarchisation et de focalisation, la prosodie assume aussi les fonctions de modalisation qui sont liées aux modalités de phrase : assertive, interrogative et impérative. Chaque modalité est caractérisée par un contour prosodique différent, par exemple, l'intonation de la phrase déclarative est généralement descendante à la fin, celle d'une question simple est plutôt montante (Delattre, 1966; Aubergé, 1992; Hirst & Di Cristo, 1998; Yuan et al., 2002).

Au-delà des fonctions linguistiques, la prosodie prend en charge aussi des fonctions expressives dans la communication parlée. Pour elle, les fonctions expressives de la prosodie varient à deux niveaux différents dans la communication interindividuelle, autrement dit, la prosodie remplit deux fonctions distinctes pour exprimer l'affect: fonction émotionnelle et fonction attitudinale. Il s'agit donc d'un modèle fonctionnel de la prosodie où trois fonctions différentes sont distinguées : fonctions linguistiques (présentées ci-dessus), fonctions attitudinales et fonctions émotionnelles. Ce qui nous intéresse plus particulièrement ici, c'est la fonction attitudinale de la prosodie. Ce modèle fonctionnel de la prosodie est illustré par la figure 2.

Selon Aubergé, la distinction essentielle réside dans les domaines temporels correspondants aux différents types de contrôle cognitif : les fonctions attitudinales, qui ajoutent aux fonctions linguistiques des valeurs sur les intentions du locuteur, sont contrôlées volontairement par le locuteur et sont régies par le « temps linguistique »; les expressions émotionnelles, qui relèvent de l'expression directe des émotions du locuteur, sont régies par un temps émotionnel et sous un contrôle involontaire. De plus, comme mentionné dans la section précédente, les attitudes et les affects sociaux sont des codes construits et acquis dans la dynamique du langage et de ses réalités socioculturelles, les réalisations prosodiques des affects sociaux peuvent aussi varier selon la langue et culture (Shochi et al., 2005; Rilliard et al., 2009).

Parallèlement à ce modèle fonctionnel de la prosodie sur lequel est basé notre travail, Aubergé (2002a, 2003) a précisé le rôle de la prosodie dans la communication interindividuelle en proposant un modèle du système de communication que nous allons présenter dans la sous-section suivante.

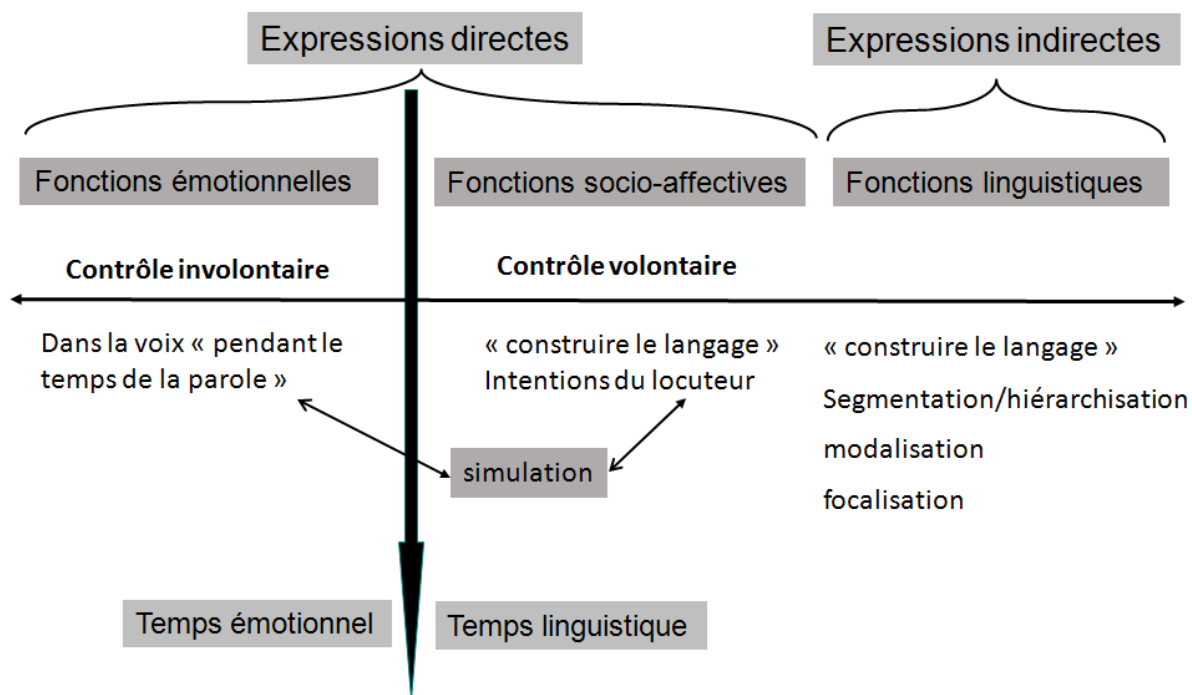


Figure 2 : Le modèle fonctionnel de la prosodie proposé par Aubergé (2002a, 2003)

#### 1.3.4. Prosodie dans le système de communication parlée: le modèle C-Clone

Le rôle de la prosodie dans la communication humaine est depuis longtemps considéré comme guidé par les fonctions syntaxiques, même congruente avec la syntaxe selon quelques auteurs, comme le rappelle par exemple Caelen-Haumont & Keller (1997) et Rossi (1999). Cependant, en réalité, la prosodie prend en charge beaucoup d'autres fonctions que la syntaxe de l'énonciation.

En intégrant son modèle prosodique, Aubergé (2002a, 2003) a proposé un autre modèle du système communicatif, appelé le modèle C-Clone, illustré dans la figure 3. C'est une architecture cognitive qui décrit en détail le processus de communication humaine durant lequel les multiples modules se mettent en synergie coopérative et s'auto-organisent en impliquant tous les matériaux dans la communication sur le principe d'un contrôle fonctionnaliste. Les modules, à savoir la phonologie, le lexique, la

morphonologie, la syntaxe, la prosodie, l'expression faciale et la gestualité, interagissent entre eux en gardant chacun en quelque sorte une autonomie selon des degrés de liberté et des contraintes qui leur sont spécifiques. Il s'agit, ici, de rendez-vous structurels (Aubergé, 1991). Dans ce modèle communicatif composé de multi-agents, la prosodie est le seul agent qui remplit toutes les fonctions en coopération avec les autres agents du système, et la fonction attitudinale de la prosodie émerge également dans l'interaction avec les mêmes modules.

En bref, le modèle d'Aubergé considère le système de communication humaine comme l'interaction d'un ensemble d'agents linguistiques autonomes, y compris la prosodie, qui s'auto-organisent en coopération. À part les fonctions linguistiques, les fonctions de l'expression des émotions et attitudes relèvent également des fonctions globales du système communicatif et sont réparties de façon interactive entre les agents du système. Cette coopération d'inter-agent est assurée par des points de rendez-vous structurels entre la prosodie et les niveaux linguistiques/paralinguistiques (Aubergé, 1991).

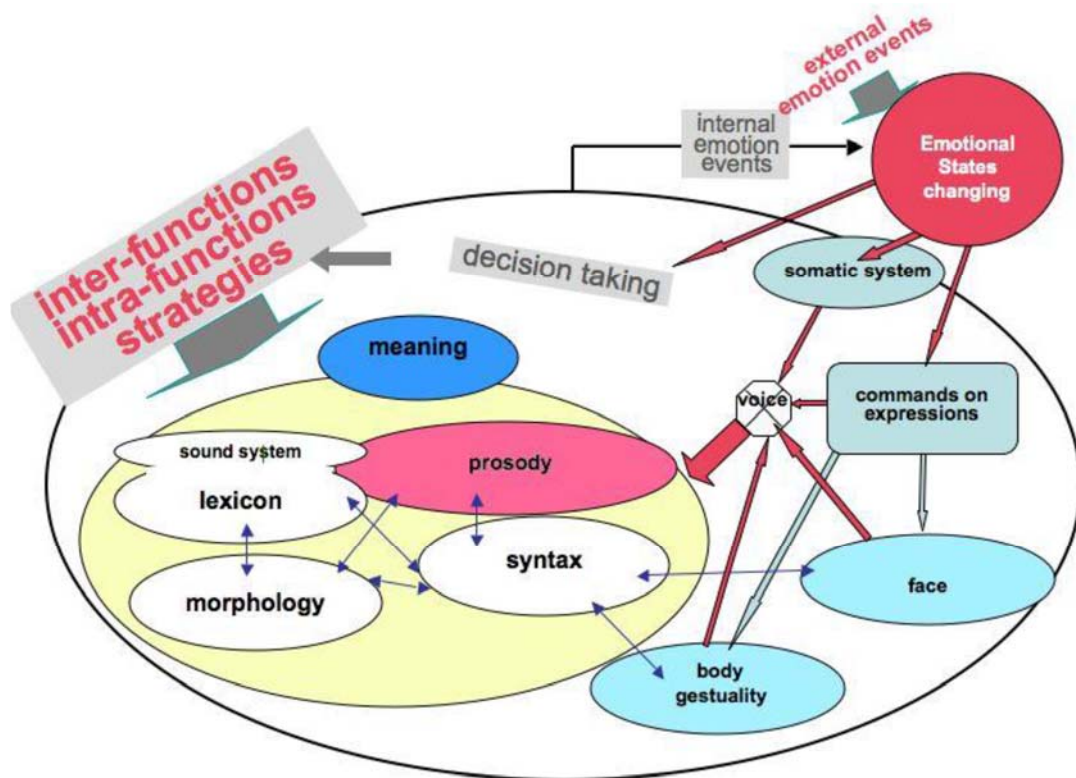


Figure 3 : Représentation du système de communication réparti sur les multi-agents en coopération interactive (Aubergé, 2003)



### 1.3.5. Quelques premières études de la didactique de la prosodie attitudinale

Concernant la relation entre la prosodie et l'attitude, il faut remonter à la fin du 19<sup>ème</sup> siècle pour trouver les premières recherches élaborées sur les attitudes du français. Passy (1890/1979) a signalé l'importance des traits uniquement mélodiques pour exprimer le sens et la signification résultant de l'émotion et l'attitude révélées par le locuteur dans son énonciation. Il en a montré quelques exemples en remarquant les différents effets des variations du contour mélodique : par exemple, une interjection comme « ah ! » prononcée avec une montée faible, indique curiosité et intérêt ; avec une montée forte, étonnement.

Dans un autre exemple, il a distingué 7 contours mélodiques différents du mot « oui », qui ont donné chacun un effet différent sur l'énonciation.

<i>Oui \</i> <sup>11</sup>		<i>[C'est mon avis]</i>
<i>Oui \</i>		<i>[J'affirme cela]</i>
<i>Oui /</i>		<i>[Est-ce vrai?]</i>
<i>Oui /</i>	<i>montée forte</i>	<i>[Pas possible!]</i>
<i>Oui \ /</i>		<i>[C'est possible mais j'en doute]</i>
<i>Oui /\</i>		<i>[C'est bien clair]</i>
<i>Oui /\ /</i> <i>mais...]</i>		<i>[Sans doute, au premier abord ;</i>

(Léon, 2000, p. 134)

Bien qu'extrêmement limités, les travaux de Passy constituent d'excellentes recherches phonostylistiques, qui marquent les débuts des recherches sur les attitudes. A son instar, d'autres chercheurs, comme D. Callamand et P. Delattre, ont posé leur regard sur les attitudes prosodiques en les considérant comme stratégie pragmatique au sein de l'acte de parole. Ce point de vue a été largement partagé parmi les didacticiens des langues.

En réalité, concernant l'acquisition de la prosodie, elle a fait depuis longtemps l'objet de nombreuses recherches chez le bébé à partir de 6 mois (Crystal, 1986 ; Dore, 1975 ; Halliday, 1975 ; Konopczynski, 2000 ; de Boysson-Bardies, 1996 ; Billières, 2002). Les

---

<sup>11</sup> Les barres obliques indiquent la direction de la mélodie.

enfants dès l'âge de 2 ou 3 ans arrivent déjà à exprimer divers actes de langage et des modalités de manière intonativement différente (Bernicot, 1992; Bassano et al., 1998). Clément (1999) a même montré que les expressions des valeurs d'attitude sont généralement acquises dans l'enfance entre 7 ans et 11 ans. Au travers de ces recherches scientifiques, on pourrait dire que la prosodie de la langue maternelle est l'élément linguistique le plus ancré dans le comportement vocal de l'individu, et qu'elle persiste et influence souvent l'apprentissage d'une langue étrangère.

A l'instar des travaux de Delattre (1966) effectués sur les dix intonations fondamentales du français, dans les années 80 du 20<sup>e</sup> siècle, des didacticiens ont mené des recherches sur l'aspect pragmatique de la prosodie dans une perspective didactique. Callamand (1991) a décrit cinq attitudes du français selon leurs traits prosodiques obtenus par l'impression auditive et leur a également attribué des valeurs communicatives en fonction de leur usage pragmatique et didactique chez les apprenants étrangers. Selon Callamand (1991, p. 6) :

- le [ton bas – ton haut – ton bas] caractérise l'ironie,
- le ton mélodique est indissociable de l'expression des sentiments,
- le ton suraigu traduit l'évidence ou l'indignation,
- le ton de justification,
- le [ton bas + chuchotement] dévoile une attitude méprisante.

Elle a aussi indiqué qu'il existe bel et bien des caractéristiques individuelles qui servent de marqueur socioculturel permettant d'identifier le locuteur et l'aider à réaliser son désir de produire un effet spécifique. Par exemple, l'allongement de la syllabe finale et la montée de la hauteur pourrait donner une impression de légèreté calculée pour la séduction ; un allongement ajouté sur la dernière syllabe du groupe manifestant un ton affecté.

Mejvaldová (2000) a effectué dans ses travaux la comparaison interculturelle de la prosodie des attitudes du français et du tchèque. Dans ses travaux, 15 phrases de modalités affirmatives et interrogatives ont été produites par des locuteurs français et tchèques qui sont tous bilingues dans les deux langues, selon 8 expressions affectives : peur, colère, tristesse, ennui, surprise, admiration, joie et neutralité. Le résultat de l'évaluation perceptive de ces stimuli auprès des sujets français et tchèques a montré

que les auditeurs ont mieux perçu les attitudes de leur langue maternelle ; les auditeurs ayant une bonne connaissance de la langue (le français pour les sujets tchèques et le tchèque pour les sujets français) avaient une facilité d'identification des attitudes transmises par la prosodie. Or, quelques potentiels « faux-amis » prosodiques ont aussi été identifiés dans les tests perceptifs, il s'agissait des attitudes qui ont été bien perçues par les auditeurs natifs mais mal reconnues par les auditeurs étrangers. Par exemple, l'attitude de l'ennui produite en français a été interprétée par les auditeurs tchèques comme « triste » ou « furieux » ; la « déclaration » et la « surprise » exprimées en tchèque ont été considérées comme « joyeux » par les sujets français.

En effet, l'impact des « faux-amis » prosodiques sur la perception des expressions affectives chez les sujets étrangers a déjà été prouvé dans diverses langues par des études interculturelles de la prosodie des attitudes (Abelin & Allwood, 2000; de Moraes et al., 2010; Mac et al., 2010; Shochi et al., 2010). Cela nous conduit à croire que l'apprentissage de la prosodie des affects sociaux de la langue cible est incontestablement nécessaire pour les apprenants en langue étrangère (LE) et seconde (L2).

Or, sous l'influence de sa langue maternelle, l'apprenant d'une LE ou L2 s'approprie souvent lors de l'apprentissage de la langue cible, une interlangue, définie comme un système de règles linguistiques abstraites propre aux apprenants LE et L2. Elle constitue un stade intermédiaire entre la langue source et la langue cible (Selinker, 1972, 1992; Ellis, 2000). L'apprenant transfère inconsciemment les caractéristiques phonético-phonologiques et prosodiques de sa langue maternelle sur la langue cible. Ceci l'empêche de parler naturellement et de façon plus appropriée la langue étrangère. Du côté de l'enseignement des langues étrangères, les manuels pédagogiques proposent en général un travail sur l'identification et la production des phonèmes en langue cible mais l'aspect strictement prosodique de la langue est souvent traité de façon superficielle, sans parler de la prosodie des affects sociaux. Ainsi, faute d'un enseignement spécifique portant sur la prosodie de la langue cible, l'apprenant, quel que soit son niveau de langue cible, a, dans la plupart des cas, tendance à segmenter le flux verbal en fonction des stratégies prosodiques acquises et développées en langue maternelle pendant son enfance. Il transfère facilement les contours prosodiques de sa langue maternelle sur la langue étrangère en oral spontané comme en lecture oralisée (Billières, 2005).

En synthétisant les études citées ci-dessus, nous pourrions en tirer la conclusion suivante : vu que la prosodie est étroitement liée à la langue et à la culture, l'enseignement/apprentissage de la prosodie d'une langue étrangère est nécessaire aux apprenants d'une LE ou L2 ; la perception et l'expression des affects sociaux en langue étrangère et seconde sont souvent influencées par la langue maternelle de l'apprenant, ce qui s'observe par des « faux-amis » prosodiques sur le plan perceptif et par des expressions intonativement inappropriées sur le plan expressif. Ceci met en évidence la nécessité de l'enseignement de la prosodie des affects sociaux en langues étrangères et secondes.

#### 1.3.6. Enseignement de la prosodie du Français Langue Etrangère en Chine

En remontant l'histoire de l'enseignement du Français Langue Etrangère (FLE) en Chine, nous pouvons constater que, depuis ses débuts au 19<sup>e</sup> siècle (Z. Li, 1990; Zhang, 1990) jusqu'à aujourd'hui, de nombreuses méthodologies et manuels ont paru et chacun a occupé successivement sa place d'une période à l'autre (Fu, 2006; Pu et al., 2006). Parmi ces méthodologies, la méthodologie traditionnelle est la méthode la plus longtemps utilisée et est toujours en vigueur dans les pratiques de classe à l'université.

Aussi nommée « méthode grammaire-traduction », la méthode traditionnelle est basée sur l'enseignement des règles grammaticales qui sont en grande partie décontextualisées, et sur des listes de mots de vocabulaire avec des exercices d'application. Aux dépens de l'oral, l'écrit prime en didactique du FLE privilégiant la compétence linguistique plutôt que la compétence de communication. Il en résulte que des apprenants qui sont meilleurs en compréhension et expression écrites, ont une prise de parole et une production orale très faible. Par ailleurs, « *lorsque les supports de cours utilisés reposent essentiellement sur l'écrit, la dimension prosodique de la parole est singulièrement réduite* » (Detey, 2005).

Cette hégémonie de la méthodologie traditionnelle des langues étrangères en Chine a pour origine de multiples raisons historique, politique, sociale, etc., mais ceci ne constitue pas notre sujet de recherche. Ce qui nous intéresse, c'est quelle place la prosodie occupe et va occuper dans l'idéologie scolaire de la société cible.

Au sujet de ce qu'est la prosodie, nous avons mené des discussions avec des enseignants de français de l'Université des Etudes Internationales de Xi'an en Chine, parmi lesquels certains très expérimentés et d'autres en début de carrière. Ces discussions nous ont permis de constater que la plupart de ces enseignants ignorent ce qu'est la prosodie ; très peu d'enseignants affirment leur connaissance, et lorsqu'ils doivent expliquer comment ils enseignent la prosodie, ils donnent des exemples plutôt de correction phonétique (paire minimale, par exemple) ou d'enseignement de l'intonation de phrase assertive et interrogative.

Un autre constat a été fait par Pernet-Liu (2013) sur les sujets de mémoires d'étudiants chinois en FLE. L'auteur a ramassé 743 mémoires écrits entre 2004 et 2012 à l'Université des Langues étrangères de Pékin (Beiwai) et à l'Université des Langues et Cultures de Pékin (Beiyu), dont les départements de français sont très réputés en Chine. Le résultat de l'analyse du corpus montre que, au niveau des thématiques, les mémoires des étudiants de Beiwai portent plus sur des sujets de questions sociales (23%), de questions politiques (18%) et de questions économiques (17%) ; ceux des étudiants de Beiyu sont plus au sujet de la linguistique (23%) (la grammaire, le lexique et la phonétique sont des sujets plus abordés en linguistique française), et des questions sociales (20%). La littérature et les thématiques culturelles viennent derrière les thématiques les plus abordées. Evidemment, aucun mémoire dans le corpus ne traite la prosodie. Le résultat de cette recherche révèle la structuration des disciplines en français dans les établissements d'enseignement supérieurs chinois où l'enseignement de la prosodie est toujours absent.

Or, du fait de l'évolution de la société et l'accroissement constant des échanges internationaux, les objectifs de la formation des apprenants en langues étrangères sont rajustés en fonction de la demande du marché de l'emploi. Les enseignants de langues étrangères prennent en charge non seulement l'enseignement des savoirs linguistiques, mais aussi des savoir-faire, pour former des apprenants dits « de compétences composées » (Fu, 2006). C'est un grand challenge pour tous les enseignants de langues étrangères, pas seulement du français. Il est également le bon moment pour eux de réfléchir sur leur travail exact.

Il n'y a pas de doute que la compétence de communication est une compétence cruciale pour les apprenants en langues étrangères. Pour former une vraie compétence de

communication en langue étrangère, il faut savoir co-construire les formes morpho-lexico-syntaxiques à l'aide de leurs réalisations pertinentes en fonction de la nature et du contexte socioculturel de l'interaction et savoir exprimer, par le biais de la prosodie en particulier, son opinion, son intention, son attitude pendant l'acte de parole.

Si l'enseignement de la prosodie du français n'a pas eu sa place dans l'histoire de l'enseignement du FLE en Chine, il méritera désormais une place importante, surtout quand nous nous demandons pourquoi nos apprenants présentent souvent dans la communication orale des lacunes importantes même s'ils acquièrent une prononciation « acceptable » en français.

## **1.4. Chinois mandarin et sa prosodie**

Après une synthèse des théories de l'émotion, de l'attitude, et de la prosodie, nous aborderons au sein de cette section la prosodie du chinois mandarin et présenterons quelques travaux effectués sur la prosodie des émotions et des attitudes en chinois mandarin.

### **1.4.1. Présentation du chinois mandarin**

#### **1.4.1.1. Chinois mandarin et les langues chinoises**

Le chinois mandarin ou le chinois moderne, également nommé « langue nationale », « langue commune », « langue des Han » (ethnie dominante en Chine), est la langue officielle en Chine continentale, à Taiwan et à Singapour. Il est aussi très souvent nommé simplement « chinois », et considéré comme la langue la plus parlée au monde, avec plus de 850 millions de locuteurs, soit 70% de la population chinoise (Yang-Drocourt, 2007). Typologiquement, il appartient au groupe sino-tibétain, divisé en deux grandes familles de langues : les langues tibéto-karens (ou karènes) et les langues Han, communément nommées langues chinoises. En Chine continentale, il y a 7 grands groupes dialectaux chinois dont le nombre de locuteurs est important : Mandarin, Cantonais, Wu, Gan, Hakka, Min et Xiang, parmi lesquels on cite toujours en premier le mandarin, appellation générale des langues parlées au Nord, à l'Est et au Sud-Ouest de la Chine continentale. Le chinois mandarin, ou le mandarin standard, est la variante standardisée du mandarin du Nord et principalement basée sur le parler de Pékin (Chao, 1968; Li & Tompson, 1981; Yuan, 2001). Par ailleurs, quelques langues non-han sont également

parlées en Chine, principalement dans les régions de Mongolie intérieure, du Tibet et des minorités ethniques au Yunnan et Guizhou. La carte de la figure 4 présente la distribution géographique des langues chinoises en Chine continentale.

Le chinois mandarin est une langue SVO, c'est-à-dire que son ordre syntaxique le plus courant est l'ordre « sujet – verbe – objet » (Sun & Givon, 1985; Y. A. Li, 1990). La morphologie n'a pas autant d'importance en chinois qu'en français : il n'y a pas de variation des noms, des adjectifs et des pronoms en fonction du genre et du nombre ; les verbes ne se conjuguent pas selon le sujet, le temps et le mode de phrase, et le temps d'une action est souvent indiqué au moyen d'un adverbe.

Le chinois mandarin possède deux systèmes différents : un système d'écriture logographique dont l'unité de base est le caractère appelé sinogramme, et un système alphabétique qui sert de la transcription phonétique officielle de la langue. Bien que l'écriture chinoise soit maintenant unifiée et standardisée, les caractères chinois sont des morphèmes logosyllabiques indépendants des changements phonétiques (Gelb, 1973), par exemple, un même caractère peut être différemment prononcé dans les différentes langues modernes de la Chine. Par conséquent, l'écriture chinoise est indépendante de la langue parlée. Pour noter la prononciation des caractères, on utilise le Pinyin qui est le système officiel de transcription du chinois mandarin. C'est un système phonétique où tous les phonèmes du chinois (consonnes, voyelles, semi-voyelles) sont transcrits en alphabet latin, plus le ton représenté par des symboles indiquant la direction de la mélodie.

Pour conclure cette brève présentation, tout d'abord au niveau de la terminologie, il faut distinguer le mandarin et le chinois mandarin : le premier indique un groupe dialectal parlée principalement au Nord de la Chine, et sur laquelle se base le chinois mandarin devenu la langue officielle de la Chine de nos jours. Dans notre travail de thèse, les termes du « chinois », de la « langue chinoise » se réfèrent tous au chinois mandarin. Par ailleurs, au niveau linguistique, le chinois mandarin est une langue composée de morphèmes indépendants, dont la frontière est identique avec celle de la syllabe, il est donc une langue analytique ou isolante. Enfin, il est aussi une langue à tons lexicaux qui sont problématiques dans l'apprentissage du chinois langue étrangère. Nous allons donner plus de détails concernant la structure syllabique et les tons lexicaux du chinois mandarin dans les sous-sections suivantes.

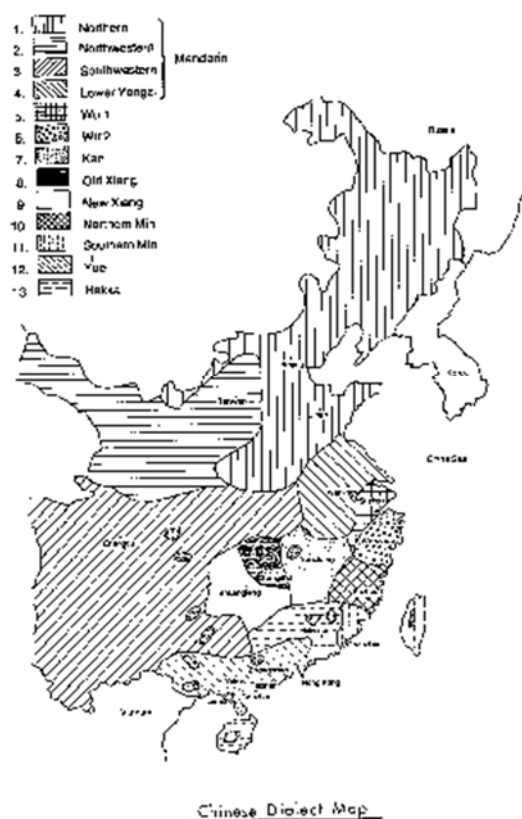


Figure 4 : Carte de la distribution géographique des langues chinoises (Li & Tompson, 1981).

#### 1.4.1.2. Structure syllabique du chinois mandarin

En chinois, la syllabe est classiquement considérée comme l'unité phonologique fondamentale, contrairement aux phonèmes dans le contexte linguistique occidental. Ce qui est une caractéristique essentielle du chinois mandarin. Cependant, cela ne signifie pas que chaque lexème est représenté par une syllabe, mais plutôt que chaque morphème correspond généralement à une syllabe entière, dont le graphème est un seul caractère (mot monosyllabique). Il est consensuellement admis que le chinois moderne ou chinois mandarin est plutôt une langue dissyllabique (Duanmu, 2002). En réalité, les mots chinois sont en longueur variable : monosyllabique, dissyllabique, trisyllabique, etc., mais selon une enquête officielle du lexique chinois mené en 1959 par l'organisation *Chinese Language Reform Committee Research and Popularization Office*, ce sont les mots dissyllabiques qui s'imposent quantitativement dans le vocabulaire moderne.



En ce qui concerne la structure syllabique du chinois mandarin, la phonologie traditionnelle du chinois considère que la structure fondamentale est une combinaison de deux unités distinctes au sein de la syllabe (Wu, 2000) : une initiale, qui correspond à une consonne simple ou nulle, et une finale, correspondant à une voyelle (monophtongue, diphtongue ou triphthongue) parfois accompagnée d'une finale nasale. Par exemple : « 马 » mǎ (cheval), est composé d'une seule syllabe dont l'attaque est /m/ et le noyau vocalique est la voyelle /a/ ; le mot « 繁忙 » fán - máng (occupé) contient deux syllabes, et la structure de chaque syllabe est identique : consonne initiale – voyelle principale – coda consonantique.

Il y a 22 consonnes et 5 voyelles fondamentales en chinois (sans compter les variantes de ces voyelles fondamentales). Cependant, il n'y a qu'environ 407 syllabes non tonalisées en chinois et a peu près 1250 si l'on inclut les tons (Deng & Dang, 2007).

#### 1.4.1.3. Tons lexicaux du chinois mandarin

Les tons lexicaux sont superposés à la syllabe qui décide le positionnement et la durée des tons. Le ton est une partie intégrante de la syllabe et lui accorde non seulement une mélodie mais aussi une signification (ce qui n'inclut pas l'homophonie). On pourrait dire que deux syllabes homophones mais de ton différent peuvent en principe être homonymes (Martin, 1990). Ce constat nous conduit à penser que les tons constituent une grande difficulté pour les apprenants qui n'en ont pas dans leur langue d'origine.

Il existe 4 tons ou 5 avec le ton neutre, qui est considéré aussi comme une absence de ton. Les quatre tons, portés par la voyelle principale, concernent toujours une syllabe dans sa totalité et ne traversent pas les frontières de la syllabe. On a coutume de noter les tons chinois comme ton 1, ton 2, ton 3 et ton 4. Le tableau ci-dessous présente une description brève des tons chinois :

Ton	Appellation chinoise	Appellation française	Appellation anglaise	Symbole	Hauteur <sup>12</sup>	Exemple
1	阴平 (yīnpíng)	Ton haut et plat	High level tone	ˉ	55	mā (ma1)
2	阳平 (yángpíng)	Ton montant	Rising tone	ˊ	35	má (ma2)
3	上声 (shàngshēng)	Ton descendant-remontant	Falling rising tone	ˇ	214	mǎ (ma3)
4	去声 (qùshēng)	Ton descendant	Falling tone	ˋ	51	mà (ma4)

Table 2 : Description des quatre tons chinois.

En ce qui concerne la représentation des tons du chinois mandarin, la proposition communément admise par les phonologues et phonéticiens du chinois est celle de (Chao, 1968), qui a suggéré de représenter les quatre tons chinois sur une échelle logarithmique à 5 niveaux : niveau 1 représente le plus bas niveau de pitch et 5 le plus haut niveau (figure 5). Cette représentation schématique n'est pas très éloignée du paramètre de F0 et représente plutôt la réalisation maximale des 4 tons lorsqu'ils sont prononcés isolément (Belotel-Grenié & Grenié, 1997).

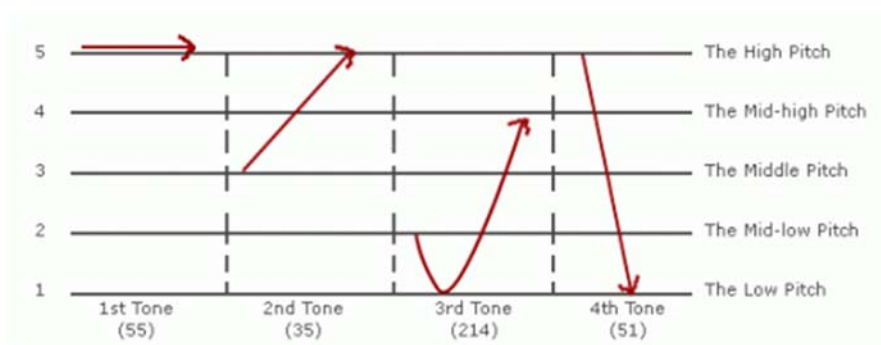


Figure 5 : Représentation schématique des quatre tons du chinois mandarin proposée par Chao (1968)  
(<http://www.chine-culture.com/chinois/pinyin.php>)

#### 1.4.2. Prosodie du chinois mandarin

Après cette vue globale sur le chinois mandarin et ses caractéristiques proéminentes, nous allons entamer dans cette sous-section les caractéristiques de la prosodie du chinois mandarin, à partir d'un point de vue fonctionnel de la prosodie, que nous avons présenté dans la partie 1.3.3.2.

<sup>12</sup> En référence à Chao (1968).

#### 1.4.2.1. Fonctions linguistiques de la prosodie du chinois mandarin

Comme toutes les langues tonales, la prosodie de la langue chinoise assume des fonctions linguistiques non seulement au niveau de l'énoncé mais aussi au niveau de la syllabe.

##### Au niveau de la syllabe

La prosodie du chinois mandarin, comme celle du vietnamien et du thaï, prend en charge une fonction tonale ou fonction lexicale distinctive (Di Cristo, 2013). Chaque ton possède un contour prosodique spécifique, et permet de distinguer la signification lexicale d'un mot.

En chinois mandarin, le ton est principalement déterminé par la fréquence fondamentale (F0) créée au niveau du système laryngé. Parmi les paramètres prosodiques (la fréquence fondamentale, l'intensité et la durée), la hauteur et le contour de F0 sont les paramètres acoustiques primaires pour caractériser les tons chinois (Dreher & Lee, 1968; Howie, 1970; Chuang et al., 1972; Moore & Jongman, 1997). La figure 6 présente une illustration des courbes de F0 des quatre tons chinois produits par une locutrice native. En général, le ton 1 est un ton haut et relativement plat sur la durée totale ; le ton 2 est un ton montant dans la plupart de sa durée, mais le début de la montée se trouve dans la zone centrale de la plage de F0 et atteint à la fin la hauteur de F0 du ton 1 ; le ton 3 est le ton le plus bas parmi les quatre tons, bien qu'il remonte jusqu'à la fin après une descente légère ; le ton 4 est un ton descendant, qui commence avec une hauteur élevée et descend ensuite jusqu'au bas de la plage de F0.

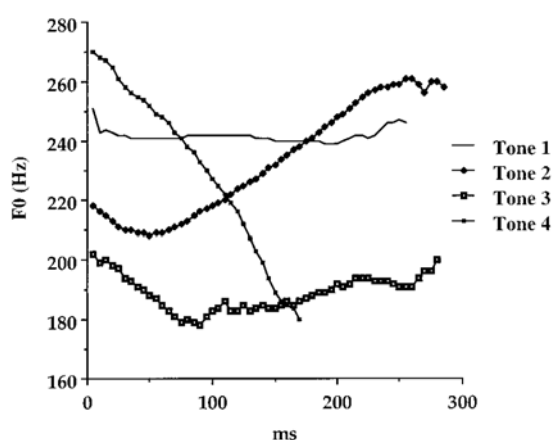


Figure 6 : Contours de F0 des quatre tons du chinois mandarin présentés dans le contexte segmental de/ma/, prononcé isolément par une locutrice native (Moore & Jongman, 1997).

Beaucoup d'études ont déjà montré que F0 est un indice essentiel pour la perception des tons du chinois mandarin (par exemple Howie, 1976), et notamment la hauteur et le contour de F0. Gandour (1984) a supposé que, bien que ces deux dimensions soient très importantes, les auditeurs chinois attachent légèrement plus d'importance au contour de F0 qu'à la hauteur de F0. Au contraire, soutenus d'une étude de la perception des tons 1 et 2, Massaro et al. (1985) ont déclaré que ni la hauteur ni le contour ne peut aider seul les auditeurs à correctement identifier les tons, mais plutôt que les deux ont été utilisés par les auditeurs.

En plus des différences en F0, les tons chinois montrent également certaines différences en termes de durée totale : les tons 2 et 3 apparaissent les plus longs et le ton 4 le plus court (Dreher & Lee, 1968; Chuang et al., 1972; Nordenhake & Svantesson, 1983). En ce qui concerne l'amplitude des quatre tons, Chuang et al. (1972) ont montré que le ton 4 a l'amplitude totale la plus forte et le ton 3 la plus faible. Bien que les tons manifestent des différences dans les trois dimensions prosodiques, les trois paramètres acoustiques n'ont pas la même importance dans la perception des tons : F0 est l'indice prédominant et la durée et l'intensité n'ont qu'un petit effet sur la perception (Lin, 1988).

A part les caractéristiques prosodiques des tons du chinois mandarin, un autre phénomène important dans la production et la perception des tons porte sur la coarticulation tonale ou « assimilation tonale » dans la parole continue (Han & Kim, 1974; Brunelle, 2003). Dans la première étude complète portant sur la coarticulation tonale en chinois, Shen (1990) a montré que la coarticulation est bidirectionnelle et influence surtout la hauteur de F0 du ton entier plutôt que son début ou sa fin. Les tons 1 et 2, qui ont une valeur de F0 élevée à la fin, ont tendance à augmenter la valeur de F0 au début du ton suivant ; le ton 4, qui a une valeur de F0 élevée au début, influence souvent le ton précédent en augmentant la valeur de F0 de sa fin. Lin & Yan (1992) ont étudié les phénomènes de coarticulation tonale dans les phrases de quatre syllabes. Contrairement à l'observation de Shen (1990), le résultat de leur analyse a montré que par rapport à la différence de la F0 intrinsèque du début et de la fin de la voyelle, la perturbation de F0 au début et à la fin de la courbe causée par la coarticulation tonale

est plus importante<sup>13</sup>. L'effet coarticulatoire est généralement unidirectionnel, et c'est toujours le ton précédent qui influence le ton suivant, de même qu'il existe certains modèles de perturbation de F0 causée par la coarticulation tonale en chinois mandarin.

#### Au niveau de l'énoncé

Au niveau de l'énoncé, la prosodie du chinois mandarin remplit une fonction de modalisation liée aux modalités de phrase : déclarative, interrogative, impérative et exclamative (Ho, 1977; Shi, 1980), parmi lesquelles l'intonation déclarative (ou assertive) et interrogative est l'intonation fondamentale de presque toutes les langues. En général, l'intonation déclarative est marquée par une mélodie descendante à la fin de phrase et l'intonation interrogative une mélodie ascendante. Toutefois, en tant qu'une langue tonale, la différence entre l'intonation des deux modalités d'énonciation est plus compliquée en chinois mandarin en raison de l'interaction entre l'intonation et le ton lexical.

Shen (1990) s'est concentré sur la différence entre l'intonation déclarative et l'intonation interrogative du chinois mandarin. Elle a enregistré un corpus avec des variations syntaxiques (3 types de structure syntaxique), des variations de tons (chaque phrase est réalisée respectivement avec les 4 tons ; dans chaque phrase, les syllabes ont le même ton) et des variations de modalité (déclaration, question totale et question partielle). Le résultat du test perceptif a montré que, premièrement, sans considérer la F0 de la fin de phrase, l'intonation interrogative débute d'un niveau de registre plus haut que l'intonation déclarative ; deuxièmement, l'intonation perturbe le ton en influençant la valeur de F0 du ton et son contour mélodique, en particulier le contour du ton 1 (ton haut et plat), qui est transformé du plateau à la pente. Cependant, ce changement ne perturbe pas la reconnaissance du ton. De plus, L'intonation interrogative augmente non seulement la valeur de F0 de la dernière syllabe de phrase, mais aussi celle de toute la phrase. Ce qui a été confirmé plus tard dans les travaux de Yuan et al. (2002). Les résultats de Shen (1990) ont confirmé les études de Ho (1976), de Tseng (1981) et de Wu (1982), qui ont aussi dévoilé que les valeurs de F0 des quatre tons sont sous l'influence de l'intonation de phrase à deux niveaux : le registre et la forme.

---

<sup>13</sup> «The F0 perturbation at the start-point and the ending-point of the F0 curves in each syllable caused by tonal coarticulation is larger than the difference of the intrinsic F0 of vowels at the start-point and the ending-point » (Lin & Yan, 1992).

Yuan (2004) s'est intéressé au même sujet mais en mettant l'accent sur l'identification de l'intonation déclarative et interrogative. Il a tiré trois conclusions principales des résultats de l'expérimentation : tout d'abord, la déclaration est plus facile à identifier que l'interrogation, et beaucoup d'énoncés interrogatifs ont été perçus comme la déclaration. Le ton de la dernière syllabe de phrase n'a pas d'effet sur l'identification de l'intonation déclarative, mais perturbe celle de l'intonation interrogative. Ensuite, l'intonation interrogative réalisée avec les phrases terminées avec le ton 4 a été mieux identifiée à l'inverse des phrases terminées avec le ton 2. De plus, le focus joue également un rôle important dans l'identification de l'intonation. Un focus à la fin de phrase facilite l'identification de l'intonation de question mais complique celle de l'intonation de déclaration.

En ce qui concerne la relation entre le contour local du ton lexical et le contour global de l'énoncé, Cao (2002) a indiqué qu'en chinois mandarin, le contour local de mot ou de combinaison de mots est lexicalement décidé (par le ton ou la combinaison de tons) et reste inchangé dans l'énoncé, mais son registre, ainsi que la plage de F0, est obligé de suivre le mouvement de la pente intonative.

Différente des auteurs cités ci-dessus, Chang (1998) s'est concentré sur les caractéristiques acoustiques et la perception des questions totales en chinois mandarin (Taiwan). Dans son étude expérimentale, l'auteur a examiné les questions totales et les questions déclaratives. Selon l'auteur, les questions totales sont marquées en position finale par une particule interrogative « ma » (« 吗 ») (au ton neutre) et les questions déclaratives sont des questions totales sans particule interrogative, se caractérisant par un registre global plus élevé, ou par une montée finale plus élevée par rapport aux déclaratives correspondantes. Lorsque l'on supprime les parties de signal correspondant à la particule dans les énoncés de question totales marquées, ils ont tendance à être perçus comme l'expression de l'étonnement, ainsi que les questions déclaratives. A partir de ces observations, l'auteur a suggéré que la montée globale ou finale de ces deux types de phrase implique plus une valeur d'étonnement qu'une valeur de demande d'information. Par ailleurs, une étude acoustique confirme également que les deux types de question totale (avec ou sans la particule interrogative) ont des courbes intonatives proches avec les phrases et les questions d'étonnement. Ce qui

signifie également qu'en chinois mandarin « *l'intonation d'étonnement et celle d'interrogation sont une seule et même entité* » (Chang, 1998, p. 68).

En plus de la fonction de modalisation, l'intonation chinoise a une caractéristique qui est la déclinaison de F0 de l'énoncé. La déclinaison de la F0 est définie comme la tendance globale de la fréquence fondamentale à baisser au cours d'un énoncé, entre une ligne supérieure reliant ses pics locaux et une ligne inférieure reliant ses vallées locales qui baissent également (t'Hart & Cohen, 1973; Ladd, 1984; Cruttenden, 1997). En chinois mandarin, de nombreuses études ont été menées sur la déclinaison de l'intonation, telles que Xu (1999), Wang et al. (2001), Lin (2002) et Wang et al. (2002), parmi lesquelles certaines ont rapporté que le contour de F0 suit une tendance à baisser dans les phrases déclaratives où les tons sont mélangés (Shen, 1985; Gårding, 1987). Wang & Lin (2003) ont signalé dans leur étude qu'en chinois mandarin tous les énoncés ne subissent pas une déclinaison, mais 80% des énoncés du corpus sont affectées en dépit de la modalité de phrase. La tendance à la hausse de quelques énoncés est principalement liée au focus final et le ton lexical.

#### 1.4.2.2. Fonction expressive de la prosodie du chinois mandarin : expressions émotionnelles et attitudinales en chinois

A part la transmission du message linguistique, la prosodie tâche par ailleurs d'exprimer les émotions et les attitudes du locuteur lors d'une communication interindividuelle. Il s'agit de la fonction expressive de la prosodie (voir partie 1.3.3.2).

Par rapport aux études de la prosodie des émotions de base en chinois, les études de la prosodie attitudinale restent encore très peu développées. Nous allons présenter successivement quelques recherches sur les émotions puis sur les attitudes en chinois mandarin.

#### Études de la prosodie des émotions en chinois mandarin

Les études de la prosodie émotionnelle en chinois se concentrent plutôt sur la réalisation acoustique des émotions de base, telles que colère, joie, peur et tristesse, ainsi que leur perception.

Zhang et al. (2006) ont acoustiquement analysé les expressions vocales des émotions de colère, peur, joie et tristesse, en comparaison avec la parole neutre. Les auteurs ont

observé la fréquence fondamentale, l'intensité et la durée de ces émotions aussi bien au niveau de l'énoncé qu'au niveau de la syllabe, ainsi que l'accent tonique dans les énoncés émotionnels. Les résultats de l'analyse acoustique ont montré que, pour l'émotion de colère, au niveau de l'énoncé, la F0, la variance de F0 et l'intensité de la colère sont les plus élevées par rapport aux autres émotions, et la durée est plus courte (Yuan et al., 2002; Pao et al., 2005). Au niveau de la syllabe, l'accent tonique est plus évident et plus susceptible d'être perçu grâce à la forte énergie associée. Pour l'émotion de joie, la F0, la variance de F0 et l'intensité sont moins élevées que celles de la colère au niveau de la phrase et l'accent tonique à la syllabe n'est pas évident. Les valeurs de F0 et d'intensité de la peur au niveau de la phrase sont encore plus basses que la joie ; l'accent tonique ne constitue pas non plus un indice pour caractériser l'émotion de peur, mais l'intensité des mots non accentués de la peur est beaucoup plus basse que celle de la joie, ce qui la fait différencier de la joie et de la colère. En comparaison avec la colère, la joie et la peur, la tristesse est caractérisée par les valeurs de F0 et d'intensité les plus basses à la fois au niveau de l'énoncé et au niveau syllabique. En général, la durée n'est pas un indice important pour distinguer acoustiquement les quatre émotions. Bien que celle de la tristesse se montre la plus longue au niveau de l'énoncé, elle ne l'est pas toujours au niveau du mot.

Yuan et al. (2002) ont suggéré dans leurs travaux qu'il y a deux dimensions dans lesquelles se réalisent acoustiquement ces quatre émotions en chinois : la phonation et la prosodie. Au travers des analyses acoustiques effectuées sur un large corpus, les auteurs ont rapporté les principales caractéristiques des quatre émotions sur les deux dimensions : la joie est essentiellement caractérisée par une F0 haute et présentant une grande variation ; la colère et la peur sont principalement réalisées par la phonation : la colère est caractérisée par une voix craquée et la peur une voix soufflée. La tristesse est réalisée à la fois en prosodie et en phonation, qui est caractérisée par une voix soufflée et une fréquence fondamentale basse avec une petite variation. Par ailleurs, les auteurs ont trouvé que l'articulation et la durée ont joué un rôle secondaire dans la réalisation acoustique de ces émotions.

Après avoir construit un large corpus acté composé d'expressions de six émotions (la colère, la joie, la peur, la tristesse, l'ironie et le dégoût) et d'expressions neutres, Wang & Li (2003) ont mis en œuvre dans cette étude la perception de ces émotions pour des



auditeurs chinois. Le résultat du test perceptif a montré que l'implication sémantique d'émotion n'a pas un grand effet sur la reconnaissance des émotions, ni l'entraînement des sujets avant le test. La tristesse est la mieux reconnue et le dégoût est le moins reconnu ; la peur et la tristesse ont été beaucoup confondues, ainsi que l'ironie et le dégoût ; les expressions de la joie révèlent un registre de F0 élevé, et le contour intonatif décline plus lentement que les énoncés neutres. Ces observations confirment le résultat de l'étude de Tao & Kang (2006), selon lequel la valeur de F0 de la joie et de la colère sont plus hautes que celle des énoncés neutres et la valeur de F0 de la tristesse et de la peur se montrent très proches.

### Études de la prosodie des attitudes en chinois mandarin

Jusqu'à présent, très peu d'études ont analysé les caractéristiques prosodiques et la perception des expressions attitudinales en chinois mandarin, qui restent toujours un terrain très peu exploité par rapport aux expressions émotionnelles.

Les premières recherches sur la prosodie attitudinale en chinois provenaient d'un projet de recherche du laboratoire phonétique de l'Institut de linguistique de l'Académie chinoise des Sciences sociales, en collaboration avec le centre de recherche d'IBM en Chine. Un corpus de parole expressive spontanée, « IBM-CASS corpus de dialogue » a été enregistré dans le cadre de ce projet. Ce corpus est composé de 15 dialogues téléphoniques avec le service client d'une société immobilière. 9 attitudes étaient concernées dans le corpus : gentillesse, excitation, gratitude, excuses, confusion, surprise, plainte, intéressé et impatience (Cheng, 2013, p. 10).

Sur la base de ce corpus de parole expressive, Li Aijun et son équipe ont concentré leur étude de l'attitude de gentillesse (Chen et al., 2004; Li & Wang, 2004; Li, 2005). Ils ont tout d'abord examiné les traits prosodiques de la gentillesse exprimée respectivement dans des phrases déclaratives et des phrase interrogatives, afin d'obtenir des paramètres importants pour la synthèse vocale de l'attitude de gentillesse.

Le résultat de l'analyse acoustique a rapporté que sur le plan de la fréquence fondamentale, le registre de F0 de l'expression de gentillesse, réalisée dans une phrase déclarative comme dans une phrase interrogative, est plus élevé que celui de l'expression neutre ; la plage de F0 est relativement plus large pour la gentillesse que pour l'expression neutre. Sur le plan de la durée, en comparaison avec les énoncés

neutres, les énoncés de gentillesse se caractérisent par une durée plus courte en dépit de la modalité de phrase.

Ensuite, Li et son équipe ont synthétisé des énoncés de la gentillesse en modifiant les paramètres de F0 et de durée sur la base de l'énoncé neutre. S'ensuivait un test perceptif des stimuli synthétisés avec des auditeurs natifs. Le résultat du test de la perception a montré que F0 a un rôle essentiel dans la synthèse de l'expression de gentillesse à partir des énoncés neutres ; la modification unique de la durée ne permet pas d'obtenir des expressions de gentillesse ; la modalité de l'énoncé exerce un effet sur la modalisation : les phrases interrogatives ont été mieux reconnues que les phrases déclaratives.

Gu et al. (2011) ont enrichi la liste d'attitudes dans leur étude en interrogeant les caractéristiques prosodiques et la perception de cinq classes d'attitudes. Dans chaque classe, deux attitudes opposées, une attitude positive et une attitude négative, sont comparées avec l'expression neutre. Les cinq classes d'attitudes<sup>14</sup> sont :

- Gentil vs. Hostile
- Poli vs. Rustre
- Sérieux vs. Badin
- Elogieux vs. Réprobateur
- Confiant vs. Incertain

Après avoir construit un corpus contrôlé de ces attitudes, les auteurs ont effectué un test perceptif avec des auditeurs natifs. Le résultat a indiqué que toutes les classes d'attitudes, sauf l'attitude élogieuse et réprobatrice, ont été mieux reconnues que les énoncés neutres, et les attitudes négatives ont été mieux reconnues que les attitudes positives. Concernant l'attitude de l'éloge et du reproche, c'est l'expression neutre qui est la mieux reconnue et l'attitude élogieuse a été mieux reconnue que l'attitude réprobatrice. Sur le plan des confusions entre les attitudes, pour toutes les classes sauf les attitudes élogieuse et réprobatrice, les attitudes positives sont plus confondues avec l'expression neutre, notamment l'attitude sérieuse et confiante. Pour l'éloge et le reproche, la confusion principale existe plutôt entre ces deux attitudes. Les auteurs l'ont expliqué par le fait qu'il existe un décalage entre le sens sémantique de l'expression

---

<sup>14</sup>« friendly vs. hostile; polite vs. rude; serious vs. joking; praising vs. blaming; confident vs. uncertain » (Gu et al., 2011).

attitudinale et la situation contextuelle. Ils ont aussi suggéré que ce genre d'attitudes dépend davantage du contexte de communication que les autres attitudes.

Par la suite, les auteurs ont approfondi leur étude avec une analyse acoustique menée principalement sur deux paramètres prosodiques des expressions attitudinales: la durée et la fréquence fondamentale (le registre et la plage de F0) au niveau de l'énoncé. Le résultat des tests statistiques effectués sur les valeurs des indices acoustiques extraites du corpus a montré que les attitudes opposées dans toutes les classes sont significativement différentes en termes de durée ; en ce qui concerne F0, « gentil » vs. « hostile » se montrent une différence significative en plage de F0, « sérieux » vs. « badin » en registre de F0, de même que « élogieux » vs. « réprobateur » et « confiant » vs. « incertain ».

### **1.5. Conclusion du chapitre**

Nous avons présenté dans cette partie de l'état de l'art de la thèse tout d'abord les grandes théories de l'émotion et de l'attitude en psychologie, et exposé ensuite la problématique de l'attitude vue par les linguistes, les phonostylistes et les pragmaticiens, à savoir la distinction entre l'attitude et l'émotion et la catégorisation des attitudes. Nous avons introduit par la suite une réflexion phonostylistique sur les affects sociaux. Ayant distingué attitudes vs. émotions comme une différence en terme de nature du contrôle de la part du locuteur (volontaire vs. involontaire), Aubergé (2002a) a proposé d'élargir la notion d'attitude de Fónagy avec celle d'affect social. Selon elle, à part les attitudes strictement définies au sens linguistique, il y a aussi des affects qui véhiculent des valeurs sociales hiérarchiques et définissent les rôles sociétaux des locuteurs pendant l'interaction. Aubergé les a nommés globalement affects sociaux.

Selon Aubergé, la prosodie réalise trois types de fonctions différentes dans le système de communication humaine: les fonctions linguistiques, émotionnelles et attitudinales. Les fonctions attitudinales de la prosodie se caractérisent par un contrôle direct et volontaire du locuteur sur le plan pragmatique et socioculturel, et prennent en charge des affects sociaux intentionnés de la part du locuteur dans une situation donnée. La prosodie des affects sociaux d'une langue donnée est acquise au cours de l'enfance au sein de la vie sociale, elle reflète par conséquent les aspects culturels de la communauté linguistique donnée.

Sur la base de ce modèle fonctionnel de la prosodie, nous pouvons envisager un paradigme des études de la prosodie d'attitudes, et plus largement d'affects sociaux, illustré par la figure 7.

Dans ce paradigme, il est supposé que la réussite de transmission des informations attitudinales entre le locuteur et l'interlocuteur dans une communication interculturelle est sous l'influence de trois facteurs : le concept d'affect social, sa réalisation prosodique et sa perception interculturelle. Sur le plan cognitif, les concepts des affects sociaux d'une langue peuvent exister ou pas dans une autre langue. Si les deux langues/cultures partagent le même concept des affects sociaux, il se peut que ces affects sociaux se réalisent acoustiquement d'une façon tout à fait différente. Sur le plan de la perception, les affects sociaux dont le concept est partagé ou spécifique à une culture peuvent être reconnus autant par les natifs que par les non-natifs, mais ils risquent aussi de ne pas être reconnus par les non natifs. En l'occurrence, ceux qui sont reconnus par les natifs et les non-natifs peuvent être des affects sociaux universels, et ceux qui sont reconnus uniquement par des natifs et sont identifiés paradoxalement par les non natifs peuvent être considérés comme des affects sociaux problématiques, qui poseront éventuellement des problèmes aux apprenants d'une langue étrangère. En ce qui concerne l'échec de la reconnaissance de certains affects sociaux, il est supposé que, soit parce qu'ils sont spécifiquement culturels et n'existent pas dans l'autre langue, soit parce qu'ils renvoient à un autre concept tout à fait différent chez les non-natifs, il s'agit de « faux-amis ». Cependant, si l'expression d'un concept d'affect social spécifiquement culturel ou universellement partagé est reconnue seulement par les natifs et n'interfère avec aucun concept dans l'autre langue, cet affect social peut être considéré comme inconnu dans cette langue.

Par rapport aux affects sociaux partagés ou inconnus, les affects sociaux mal reconnus par les tenants d'une autre langue apparaissent plus importants pour l'apprentissage de langue cible. Par ailleurs, la méthodologie interculturelle sera une façon convenable pour identifier le statut des affects sociaux, à savoir partagés, inconnus ou « faux-amis » (Abelin & Allwood, 2000; Scherer et al., 2001; Shochi et al., 2009), et nous servira au cours de toute notre recherche sur les affects sociaux en chinois et en français pour l'apprentissage des langues étrangères.

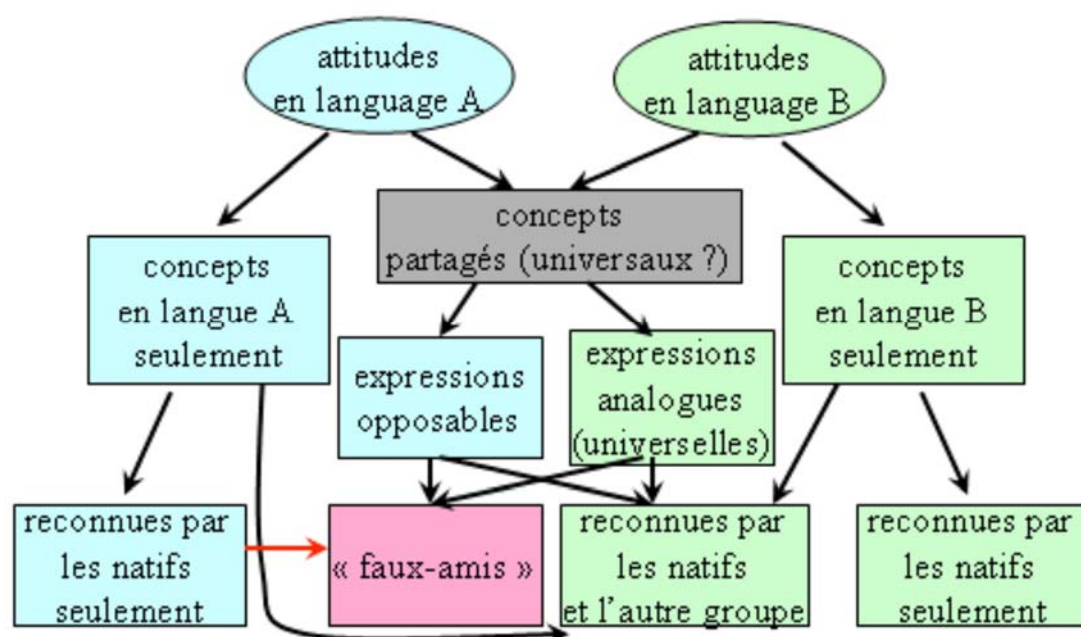


Figure 7 : Paradigme des études sur les attitudes selon le modèle d'Aubergé (Shochi, 2008, p. 43).

## **Chapitre 2. Etude intra-culturelle de la prosodie des affects sociaux chinois**

### **2.1. Introduction**

Nous nous attachons au sein de ce chapitre à une étude intra-culturelle des expressions prosodiques socio-affectives en chinois mandarin. Rappelons que nos travaux de thèse sont orientés à long terme vers l'apprentissage de la prosodie des affects sociaux du français par des chinois, et ensuite à l'apprentissage du chinois par des français. Avant d'étudier comment les apprenants perçoivent la prosodie des affects sociaux d'une langue étrangère ou seconde, nous proposons d'étudier dans un premier temps de quelle manière ils perçoivent et interprètent la prosodie socio-affective dans leur langue maternelle, et comment les natifs de la langue cible la perçoivent et comprennent eux-mêmes, et par la suite de mesurer les distances perceptives entre les apprenants et les natifs de la langue cible. C'est parce que l'apprenant projette souvent inconsciemment les expressions prosodiques de sa langue maternelle sur ses énoncés produits en langue étrangère ou seconde, même s'il suit un apprentissage spécifique de la prosodie socio-affective de la langue cible.

Ainsi, nous élaborons d'abord une étude intra-culturelle qui est d'un côté une évaluation quantifiée de la pertinence des socio-affects ainsi que de leur étiquetage et de leur méthode de production, de l'autre côté, une référence pour des études contrastives interculturelles. Pour ce faire, nous avons établi un large corpus audio-visuel des affects sociaux en chinois mandarin, variant pour chaque socio-affect en longueur syllabique et syntagmatique, ainsi qu'en valeur et position tonale. Ce corpus est ensuite évalué auditivement par des sujets chinois natifs.

Ensuite, une analyse acoustique succincte de la prosodie des expressions socio-affectives du chinois est réalisée d'abord sur la base des trois paramètres prosodiques classiques que sont la fréquence fondamentale, l'intensité et la durée, et puis sur les contours intonatifs des énoncés. Les résultats de l'analyse de la prosodie attitudinale du

chinois ont été comparés avec ceux de l'expérience perceptive afin d'interpréter les variations acoustiques à travers leur performance perceptive.

Enfin, un test perceptif des affects sociaux audio-visuels en chinois par des natifs a été mené pour examiner la contribution de la modalité auditive et de la modalité visuelle dans la perception des expressions socio-affectives.

## **2.2. Construction du corpus audio-visuel des affects sociaux en chinois mandarin**

En général, on peut opposer deux façons de recueillir des corpus expressifs. Tout d'abord, enregistrer le sujet dans la vie quotidienne (Campbell, 2004) ou lors de dialogues libres dans des contextes événementiels, soit entre humains (Douglas-Cowie et al., 2003; Devillers & Vidrascu, 2006), soit entre humain et machine (Batliner et al., 2005; Antunes et al., 2014). Il s'agit alors de corpus à la spontanéité et au naturel maximisés. Une autre approche consiste à capturer des expressions affectives soit induites par des scénarios (Banse & Scherer, 1996; Scherer et al., 2001) soit élicitées sous certaines conditions expérimentales (Scherer et al., 1984; Shochi, 2008), produites par un locuteur qui peut être un acteur professionnel ou non. Il s'agit alors de corpus actés. L'utilisation de tel ou tel type de corpus dépend de l'objectif de la recherche. Vu que notre but de recherche ne porte pas sur les expressions émotionnelles naturelles, mais sur des prototypes d'affects sociaux que l'on souhaite facilement identifiables et accessibles pour l'apprentissage des langues étrangères ou secondes, un corpus acté et enregistré en condition de laboratoire nous paraît plus pertinent qu'un corpus spontané. Par conséquent, au lieu de collecter des données spontanées, nous avons conçu et enregistré un corpus contrôlé pour 19 affects sociaux représentatifs et communs aux langues chinoise et française. Toutes les expressions enregistrées sont produites de façon prototypique et même stéréotypée par une enseignante chinoise de français qui est habituée à produire des expressions d'attitude en classe.

### **2.2.1. Choix des affects sociaux**

A l'instar de Fónagy et al. (1983), des étiquettes naïves ont été utilisées dans des études sur les attitudes françaises (par exemple : la plainte, la coquetterie, la menace, la gentillesse). Ce sont des étiquettes empiriques mais convenables à l'usage pragmatique

et vernaculaire. D'autres chercheurs en parole expressive des attitudes ont également enrichi la liste de Fónagy avec des étiquettes d'attitudes en différentes langues, par exemple : Aubergé et al. (1997) pour le français ; Diaferia (2002) pour l'anglais britannique, Shochi (2008) pour le japonais et Mac (2012) pour le vietnamien. Dans le droit fil des recherches inter-langue sur les affects sociaux, nous avons sélectionné 19 affects sociaux qui sont couramment utilisés dans la vie quotidienne aussi bien pour les chinois que pour les français. Ces étiquettes d'affects sociaux sont précisément définies et décrites dans une situation communicative, ce qui permet d'éviter les biais causés par la traduction sémantique et d'éviter de possibles mauvaises intuitions introduites par une étiquette d'affect social. Voici les étiquettes d'affects sociaux en chinois, français et anglais, ainsi que leurs définitions et leurs descriptions en contextes communicatifs.

- 陈述 (Déclaration / Declaration)

Le locuteur donne l'information sans exprimer son point de vue. Par exemple, quelqu'un lui demande l'heure, et il répond en lui donnant simplement l'information demandée : « Il est neuf heures ».

- 提问 (Question / Question)

Le locuteur pose une question, uniquement dans le but d'obtenir une information. Par exemple, quand il veut aller à un endroit qu'il ne connaît pas, il demande le chemin à un passant sans exprimer ses intentions ni ses émotions.

- 赞赏 (Admiration / Admiration)

Le locuteur exprime à son interlocuteur toute son admiration pour quelque chose, une action, un événement, un objet, ou quelqu'un, par exemple une personne célèbre ou qu'il connaît.

- 嘲讽 (Ironie / Irony)

Le locuteur répète ce que vient de lui dire son interlocuteur, mais en montrant avec humour que lui-même pense exactement le contraire de ce qu'il est en train de répéter. Cette contradiction, entre ce qui est dit et ce qu'il pense est exprimé dans sa manière de parler. Par exemple le locuteur dit : « Oui, c'est exactement ça » tout en signalant par son ton ironique que bien entendu « ce n'est pas ça du tout ».



- 蔑视 (Mépris / Contempt)

Le locuteur exprime son mépris et montre de l'arrogance vis-à-vis de ce que son interlocuteur vient d'énoncer. Par exemple il montre dans sa manière de parler qu'il considère comme sans intérêt ce qui a été dit et cherche à rabaisser son interlocuteur.

- 发怒 (Irritation / Irritation)

Le locuteur est irrité par ce qu'a fait ou dit son interlocuteur et il exprime cette irritation à son interlocuteur afin de le lui signifier clairement.

- 自信 (Confiance / Confidence)

Le locuteur a une grande confiance dans la véracité du fait qu'il énonce, il n'a aucune incertitude ni inquiétude et le manifeste dans sa manière de parler.

- 确信 (Evidence / Obviousness)

Le locuteur est sûr du fait qu'il énonce, comme dans la confiance, mais ce qu'il exprime ici à son interlocuteur, c'est que lui aussi devrait trouver certain ce fait qu'il est en train de lui énoncer. Il exprime l'évidence du fait que son interlocuteur ne devrait pas ignorer.

- 妥协 (Résignation / Resignation)

Le locuteur n'est pas convaincu par une idée ou une action que lui a proposé son interlocuteur, pourtant il doit admettre et accepter cette idée ou cette action : il énonce donc cette idée ou cette action mais exprime dans sa manière de parler le fait qu'il accepte cette idée ou cette action aux dépens de sa conviction contraire.

- 怀疑 (Doute / Doubt)

Le locuteur exprime son incertitude ou son doute par rapport à ce que vient de lui dire son interlocuteur, en reprenant dans son énoncé l'information que lui a donnée son interlocuteur. Par exemple, l'interlocuteur vient de dire au locuteur: « il va pleuvoir demain ». Mais le locuteur n'en est pas sûr car il a vu des annonces météorologiques disant qu'il pleuvra et d'autres disant qu'il fera beau demain.

- 失望 (Déception / Disappointment)

Le locuteur s'attendait avec plaisir ou espoir à un certain fait, mais quand il vient d'apprendre que ce fait ne s'est pas produit, il énonce ce fait en signalant sa déception. Par exemple, quand il voit que son équipe de foot favorite a perdu un match, il est déçu et l'exprime.

- 吃惊 (Exclamation de surprise neutre / Exclamation of neutral surprise)

Le locuteur exprime son étonnement sur le fait à propos duquel il parle et qu'il vient d'apprendre. Le locuteur est surpris par ce fait, mais il n'exprime que sa surprise, et ne donne pas son évaluation positive ou négative de cette surprise pour lui.

- 惊喜 (Exclamation de surprise positive / Exclamation of positive surprise)

Le locuteur exprime son étonnement sur un fait inattendu qu'il vient d'apprendre et qui est très positif pour lui. Par exemple, il énonce qu'il vient de gagner au loto en exprimant surprise et satisfaction.

- 惊愕 (Exclamation de surprise négative / Exclamation of negative surprise)

Le locuteur exprime son étonnement sur un fait inattendu qu'il vient d'apprendre et qui est très négatif pour lui. Par exemple, il découvre qu'on vient de lui voler son portefeuille et exprime ce fait avec surprise et consternation.

- 礼貌 (Politesse / Politeness)

Le locuteur montre de la politesse, du respect, de la courtoisie à son interlocuteur.

- 权威 (Autorité / Authority)

Le locuteur est dans une relation socio-culturelle qui lui permet d'imposer son avis à son interlocuteur en utilisant cette situation socio-hiérarchique. Cette situation peut arriver entre un responsable et son employé, ou entre un maître et son élève.

- 引诱 (Séduction / Seduction)

Le locuteur souhaite attirer l'attention de son interlocuteur en parlant, parce que son interlocuteur lui plaît et qu'il manifeste son intention de souhaiter lui plaire également.

- 儿向 (Maternel / Infant-directed speech)

C'est l'attitude propre et naturelle d'un père ou d'une mère parlant avec son très jeune enfant.

- 亲密 (Intimité / Intimacy)

Le locuteur est dans une relation de proximité sociale (amitié, famille...) avec son interlocuteur, et est dans une situation où il peut et souhaite exprimer la nature intime de cette relation.

Le tableau ci-dessous résume les 19 étiquettes d'affects sociaux en chinois et en français, ainsi que leurs abréviations. Nous utilisons l'abréviation anglaise de ces termes d'affects sociaux pour légender tous nos graphiques dans l'ensemble de nos travaux.

N°	Affects sociaux		Abréviation
	Chinois	Français	
1	陈述	Déclaration	DECL
2	提问	Question	QUES
3	赞赏	Admiration	ADMI
4	嘲讽	Ironie	IRON
5	蔑视	Mépris	CONT
6	发怒	Irritation	IRRI
7	自信	Confiance	CONF
8	妥协	Résignation	RESI
9	怀疑	Doute	DOUB
10	失望	Déception	DISA
11	确信	Evidence	OBVI
12	吃惊	Exclamation de surprise neutre	NEU-S
13	惊喜	Exclamation de surprise positive	POS-S
14	惊愕	Exclamation de surprise négative	NEG-S
15	礼貌	Politesse	POLI
16	权威	Autorité	AUTH
17	引诱	Séduction	SEDU
18	儿向	Maternel	IDS
19	亲密	Intimité	INTI

Table 3 : Affects sociaux en chinois et en français, ainsi que leur abréviation utilisée dans les travaux de thèse.

### 2.2.2. Composition du corpus

Les variations prosodiques suivantes sont prises en compte dans la conception de ce corpus d'expressions affectives: la longueur des énoncés (en syllabes), le placement des tons et la structure syntaxique des énoncés. Ces facteurs sont systématiquement variés et les valeurs prises par ces facteurs sont également systématiquement croisées entre elles, afin d'analyser l'influence de chaque facteur ultérieurement. Etant donnée la complexité de la syntaxe de la langue chinoise, la fonction syntaxique de la prosodie n'est pas abordée dans les travaux présents. Par contre, puisqu'il est impossible d'empêcher la production de la prosodie de démarcation/hiérarchisation syntaxique, nous avons pris soin de faire varier la complexité syntaxique et syntagmatique des énoncés afin d'« annuler » leur influence sur les contours prosodiques socio-affectifs. Quant à la fonction de focalisation de la prosodie, nous étions vigilants à ce que la locutrice ne la produise en aucun endroit dans son interprétation des différents énoncés. Or, ces deux fonctions de la prosodie pourraient bien sûr être une piste de recherche dans des travaux futurs. Par conséquent, nous avons conçu des phrases dont la structure syntaxique varie sur la dimension paradigmatic et syntagmatic : l'unité lexicale (mot et groupe de mots) et phrase canonique comprenant un sujet, un prédicat avec ou sans complément d'objet. Les phrases construites n'évoquent ni ne bloquent l'expression prosodique de socio-affect, à savoir qu'elles ne comprennent aucun mot qui implique un certain affect social ou une certaine émotion, et elles pourraient advenir dans tous les contextes étudiés. La table 4 donne un exemple de ce corpus. Les chiffres dans la colonne Tons représentent la combinaison de tons de l'énoncé : 1 fait référence à un ton plat, 2 à un ton montant, 3 à un ton descendant-remontant, 4 à un ton descendant et 0 à un ton neutre.

A partir de la méthodologie présentée ci-dessus, 152 énoncés sont finalement construits pour le corpus :

- Huit énoncés d'une syllabe, qui correspondent aux quatre représentations des tons chinois. Quatre énoncés sont composés de différentes structures phonémiques et les autres de la même structure phonémique.

- Vingt énoncés de deux syllabes, qui correspondent à toutes les combinaisons des tons, y compris le ton neutre à la position finale de syllabe. Ces compositions ont pour but de couvrir toutes les possibilités du phénomène de coarticulation tonale.

- Soixante-deux énoncés de trois syllabes, dont la plupart sont constituées de mots trisyllabiques. Les énoncés qui se terminent par un ton neutre sont des groupes verbaux. Les quatre tons apparaissent systématiquement à chaque position syllabique dans l'énoncé.

- Quarante-deux énoncés de quatre syllabes, dont la syntaxe se compose d'un syntagme nominal (SN) sujet, suivi d'un syntagme verbal (SV). Les frontières de syntagme varient comme ceci : syntagme nominal d'une seule syllabe suivi par un syntagme verbal de trois syllabes (SN (1) + SV (3)) ; syntagme nominal de deux syllabes suivi par un syntagme verbal d'autant de syllabes (SN (2) + SV (2)) ; syntagme nominal de trois syllabes suivi par un syntagme verbal d'une seule syllabe (SN (3) + SV (1)) ; groupe nominal composé de quatre syllabes, noté comme « SN (4) + SV (0) ». Les tons différents sont systématiquement positionnés sur la première et la dernière syllabe de l'énoncé, tandis que toutes les syllabes du milieu portent un ton plat.

- Vingt phrases de neuf syllabes, dont la syntaxe se compose d'un syntagme nominal sujet suivi d'un syntagme verbal et d'un syntagme nominal objet. Les frontières de syntagme varient en attribuant l'importance principale respectivement au sujet, au prédicat et au COD : SN(2) + SV(1) + SN(6) ; SN(7) + SV(1) + SN(1) ; SN(4) + SV(1) + SN(4) ; SN(3) + SV(3) + SN(3). Au niveau des tons, les syllabes initiale et finale des phrases portent les tons plat, montant et descendant ; le reste des syllabes porte toujours le ton plat.

Nb de syllabe	Ton	Phrase en chinois	Transcription pinyin	Traduction française	Structure syntaxique
1	1	书	shu1	livre	Mot
	2	熟	shu2	cuit	Mot
	3	数	shu3	compter	Mot
	4	树	shu4	arbre	Mot
2	1-1	声音	sheng1yin1	voix	Mot
	1-2	春节	chun1jie2	fête du printemps	Mot
	1-3	歌手	ge1shou3	chanteur	Mot
	1-4	三万	san1wan4	30 000	Mot
	1-0	他们	ta1men0	ils	Mot
3	2-1-1	直升机	zhi2sheng1ji1	hélicoptère	GN
	2-2-1	床头灯	chuang2tou2deng1	lampe de chevet	GN
4	1-1-1-0	他搬家了	ta1ban1jia1le0	Il a déménagé.	SN(1)+SV(3)
	1-1-1-2	张医生来	zhang1yi1sheng1lai2	Docteur Zhang vient.	SN(3)+SV(1)
9	4-1-1-1-1-1-1-2	四哥搬冰箱跟微波炉	si4ge1ban1bing1xiang1g en1wei1bo1lu2	Sige porte le frigo et le micro-onde.	SN(2) + SV(1) + SN(6)
	2-1-1-1-1-1-1-4	王医生他三姑妈休假	wang2yi1sheng1ta1san1 gu1ma1xiu1jia4	La troisième tante paternelle de Docteur Wang prend un congé.	SN(7) + SV(1) + SN(1)

Table 4: Exemple du corpus des affects sociaux chinois, basé sur 152 énoncés des variations de la longueur, du placement des tons et de la structure syntaxique.

### 2.2.3. Enregistrement du corpus

#### 2.2.3.1. Locuteur

Tenant compte de la finalité de ce corpus, qui est au service des études de l'apprentissage des affects sociaux prosodiques en langue étrangère ou seconde, au lieu de prendre un acteur comme locuteur, nous avons décidé d'enregistrer le corpus avec un enseignant de langue étrangère qui soit habitué à produire des énoncés réalistes pour un usage courant. Ce type de locuteur est considéré idéal par Aubergé (2002a). Sur la base de ce critère, une enseignante de français langue étrangère, de langue maternelle chinoise et originaire de la province du Shaanxi de Chine, a été choisie comme locutrice du corpus. Elle parle un chinois standard sans accent.

#### 2.2.3.2. Enregistrement

L'enregistrement a été effectué dans une chambre sourde du GIPSA-Lab. Le corpus a été enregistré à la fois au format audio et vidéo. La locutrice s'est placée à l'intérieur de la chambre devant un microphone « AKG 1000CS » fixé sur une suspension. Le microphone a été connecté à un dispositif audio qui numérisait le son à 44100Hz, 16 bits et le transmettait à un ordinateur situé à l'extérieur de la chambre sourde, sur lequel sont sauvegardés les enregistrements sonores. La gestuelle faciale de la locutrice a été enregistrée par une caméra numérique « Sony DXC-900 », qui était posée derrière le microphone et cadrant la face et le haut du buste de la locutrice. Un écran placé juste à côté de la caméra et connecté à un autre ordinateur dans la pièce extérieure affichait les énoncés à produire ainsi que l'affect social voulu. Les énoncés à oraliser ont été présentés dans l'ordre croissant des longueurs de phrases pour chaque affect social.

Étant donnée la difficulté de produire ces affects sociaux en dehors de tout contexte, une situation de communication a été imaginée pour chaque affect social, afin d'aider la locutrice à la produire le plus naturellement possible.

Un locuteur chinois natif, originaire de la même région que la locutrice prenait en charge la projection des diapositives. Il contrôlait en même temps, avec une spécialiste de la parole expressive, la performance de la locutrice pendant le déroulement de l'enregistrement. Ils observaient la locutrice en la voyant et l'entendant par le biais d'un système vidéo, et ils se sont permis d'arrêter la locutrice quand ils trouvaient qu'un énoncé n'était pas naturel, ainsi que de le lui faire reproduire. La répétition jugée la meilleure à la fois par les surveillants et la locutrice a été sélectionnée pour cet énoncé.

Le corpus final est composé de 2888 stimuli (152 énoncés \* 19 affects sociaux), pour une durée totale de 150 minutes environ. Les signaux audiovisuels ont été par la suite découpés semi-automatiquement au niveau des phrases à l'aide du logiciel «VirtualDub».

## **2.3. Perception intra-culturelle de la prosodie des affects sociaux chinois**

### **2.3.1. Introduction**

A la suite de l'élaboration du corpus audiovisuel des affects sociaux chinois, nous avons effectué une expérience perceptive pour ces affects sociaux présentés en audio seul, afin de valider le corpus et d'examiner le comportement perceptif des sujets natifs pour ces affects sociaux, ainsi que l'identification et les confusions entre affects sociaux.

Les résultats de cette expérience permettent donc d'établir une référence pour la suite des études sur les affects sociaux. Nous avons également cherché avec cette étude à vérifier si le genre, à la fois inné et stéréotypé culturellement, influence ou structure la perception des affects sociaux.

### **2.3.2. Sous-corpus**

Puisque que le corpus complet comprend 2888 stimuli soit 152 énoncés produits avec 19 affects sociaux, il n'est pas possible de valider tout le corpus avec une seule expérience. Nous avons donc sélectionné pour le test perceptif 399 stimuli (21 énoncés \* 19 affects sociaux) sur des critères de longueur, de placement des tons et de complexité syntaxique, qui constituent les paramètres de ce corpus.

Cette sélection contient quatre mots monosyllabiques, sept mots bi-syllabiques, six énoncés de quatre syllabes et quatre énoncés de neuf syllabes. Les phrases de quatre syllabes comprennent trois structures syntaxiques : un groupe nominal ; un SN sujet de trois syllabes suivi d'un verbe d'une seule syllabe ; un nom sujet d'une syllabe suivi d'un SV de trois syllabes. Les phrases de neuf syllabes contiennent deux structures syntaxiques : (1) un SN sujet de sept syllabes suivi d'un verbe d'une syllabe et d'un nom d'une syllabe ; et (2) un ensemble sujet, prédicat et complément d'objet de trois syllabes chacun. Tous les tons sont utilisés à chaque position des énoncés de deux syllabes ; seuls les tons montant et descendant sont utilisés sur la dernière syllabe des phrases de quatre syllabes et des phrases de neuf syllabes. Les autres syllabes portent toujours un ton plat. La table 5 présente les énoncés sélectionnés pour le sous-corpus.



Nb de syllabe	Ton	Phrase en chinois	Transcription pinyin	Traduction française	Structure syntaxique
1	1	书	shu1	livre	Mot
	2	熟	shu2	cuit	Mot
	3	数	shu3	compter	Mot
	4	树	shu4	arbre	Mot
2	1-1	声音	sheng1yin1	voix	Mot
	1-3	歌手	ge1shou3	chanteur	Mot
	2-2	诚实	cheng2shi2	honnête	Mot
	2-4	营业	ying2ye4	affaires	Mot
	3-1	纸箱	zhi3xiang1	carton	Mot
	4-2	放学	fang4xue2	L'école est finie	Mot
	4-4	世界	shi4jie4	monde	Mot
4	1-1-1-2	交通规则	jiao1tong1gui1ze2	Le code routier	GN
	1-1-1-2	张医生来	zhang1yi1sheng1lai2	Docteur Zhang vient.	SN(3)+SV(1)
	1-1-1-2	他刚说完	ta1gang1shuo1wan2	Il vient de terminer son discours.	SN(1)+SV(3)
	1-1-1-4	公交车站	gong1jiao1che1zhan4	l'arrêt de l'autobus	GN
	1-1-1-4	张医生唱	zhang1yi1sheng1chan g4	Docteur Zhang chante.	SN(3)+SV(1)
	1-1-1-4	他刚出去	ta1gang1chu1qu4	Il vient de sortir.	SN(1)+SV(3)
9	2-1-1-1-1-1-1-2	王医生他三姑妈出钱	wang2yi1sheng1ta1sa n1gu1ma1chu1qian2	La troisième tante paternelle de Docteur Wang paie.	SN(7) + SV(1) + SN(1)
	2-1-1-1-1-1-1-2	王医生帮她搬微波炉	wang2yi1sheng1bang 1ta1ban1wei1bo1lu2	Docteur Wang porte pour elle le micro-onde.	SN(3) + SV(3) + SN(3)
	2-1-1-1-1-1-1-4	王医生他三姑妈休假	wang2yi1sheng1ta1sa n1gu1ma1xiu1jia4	La troisième tante paternelle de Docteur Wang prend un congé.	SN(7) + SV(1) + SN(1)
	2-1-1-1-1-1-1-4	王医生帮她搬新书架	wang2yi1sheng1bang 1ta1ban1xin1shu1jia4	Docteur Wang porte pour elle la nouvelle bibliothèque.	SN(3) + SV(3) + SN(3)

Table 5 : Sous-corpus utilisé dans l'expérience perceptive intra-culturelle pour les 19 affects sociaux chinois auprès de sujets natifs.

### 2.3.3. Sujets

30 auditeurs chinois natifs (15 hommes et 15 femmes), d'âge moyen de 25,2 ans, ont participé à cette expérience. Ils sont originaires de différentes régions de Chine et ont vécu en France depuis au moins un an. La plupart d'entre eux sont étudiants à

l'Université de Grenoble. Aucun n'est allé à l'étranger avant de venir en France ni a signalé de trouble auditif.

#### 2.3.4. Protocole expérimental

Le test perceptif s'est déroulé via une interface informatique, réalisée avec le logiciel « LiveCode ». Lors de la passation, le sujet était assis confortablement face à l'écran d'ordinateur dans un lieu calme, et tous les stimuli lui étaient présentés à travers un casque audio de haute qualité. Le test débutait par une présentation globale et une description de chaque affect social illustré par des exemples de contextes possibles. Ces informations étaient écrites dans la langue maternelle du sujet. Les sujets conservaient ce document comme référence pendant le déroulement de test. Après avoir lu les consignes, les sujets commençaient le test. Chaque stimulus était présenté une seule fois, dans un ordre aléatoire. Pour chacun, les sujets doivent juger quel est l'affect social exprimé par la locutrice en choisissant une étiquette parmi les 19 proposées à l'écran. Aucune limitation de temps de réflexion avant la présentation du stimulus suivant n'est imposée. Les réponses données ont été sauvegardées automatiquement sur le disque dur de l'ordinateur. Quand le test se termine, un entretien entre l'expérimentateur et le sujet a été mené à propos de leur ressenti pendant le déroulement du test.

#### 2.3.5. Analyse des résultats

##### 2.3.5.1. Facteurs influant la reconnaissance des affects sociaux

En premier lieu, une analyse de variance (ANOVA) à trois facteurs aléatoires a été réalisée sur les réponses de sujets. L'objectif de cette analyse vise à tester la significativité des facteurs de variations sur la qualité des scores de reconnaissance des sujets. Les facteurs fixés sont le genre des sujets (Genre, 2 niveaux), l'affect social présenté (Affect social, 19 niveaux) et la longueur des phrases (Longueur, 4 niveaux). Chaque cellule de ce plan contient au moins 60 observations. Le niveau de significativité a été fixé à 0,05. L'effet du ton n'a pas été testé dans le cadre de l'expérience présente vu la complexité de modèle et également pour réduire le bruit dans l'analyse. Par contre, cet effet sera mesuré ultérieurement dans l'étude des affects sociaux chinois perçus par des non-natifs.

Le résultat de l'ANOVA est rapporté dans la table 6 où les trois facteurs, ainsi que l'interaction entre l'affect social et la longueur, et entre l'affect social et le genre, montrent un effet significatif ( $p < 0,05$ ). L'effet significatif de l'affect social explique la plus grande part de la variance observée ( $\eta^2 = 0,693$ ), avec l'interaction entre les facteurs d'affect social et de longueur ( $\eta^2 = 0,220$ ). Les facteurs de genre et de longueur ainsi que l'interaction entre l'affect social et le genre sont significatifs, mais n'expliquent qu'une part mineure de la variance. L'interaction entre les facteurs de genre et de longueur, de même que celle entre les facteurs d'affect social, de genre et de longueur, n'ont pas montré d'effet significatif ( $p > 0,05$ ).

	SCE	ddl	F	<i>p</i>	$\eta^2$
<b>Affect social</b>	<b>253,43</b>	<b>18</b>	<b>86,22</b>	<b>0,0000</b>	<b>0,693</b>
<b>Genre</b>	<b>1,97</b>	<b>1</b>	<b>12,06</b>	<b>0,0005</b>	<b>0,005</b>
<b>Longueur</b>	<b>16,19</b>	<b>3</b>	<b>33,06</b>	<b>0,0000</b>	<b>0,044</b>
<b>Affect social : Genre</b>	<b>5,57</b>	<b>18</b>	<b>1,90</b>	<b>0,0122</b>	<b>0,015</b>
<b>Affect social : Longueur</b>	<b>80,30</b>	<b>54</b>	<b>9,11</b>	<b>0,0000</b>	<b>0,220</b>
Genre : Longueur	0,13	3	0,26	0,8574	0,000
Affect social: Genre: Longueur	7,87	54	0,8929	0,6958	0,021

Table 6 : Effets des facteurs *Affect social*, *Genre*, *Longueur* et de leurs interactions sur la reconnaissance des affects sociaux chinois par des auditeurs natifs par un test ANOVA (effets significatifs en gras).

#### 2.3.5.2. Taux d'identification des affects sociaux

La figure 8 représente les taux de reconnaissance de chaque affect social. On peut y constater que presque tous les affects sociaux sont reconnus au dessus du hasard, sauf la *confiance* (CONF) (3%). La *déclaration* (DECL) (58%) et la *déception* (DISA) (58%) sont les affects sociaux les mieux reconnus parmi les 19 et avec des scores clairement supérieurs par rapport au reste des affects sociaux.

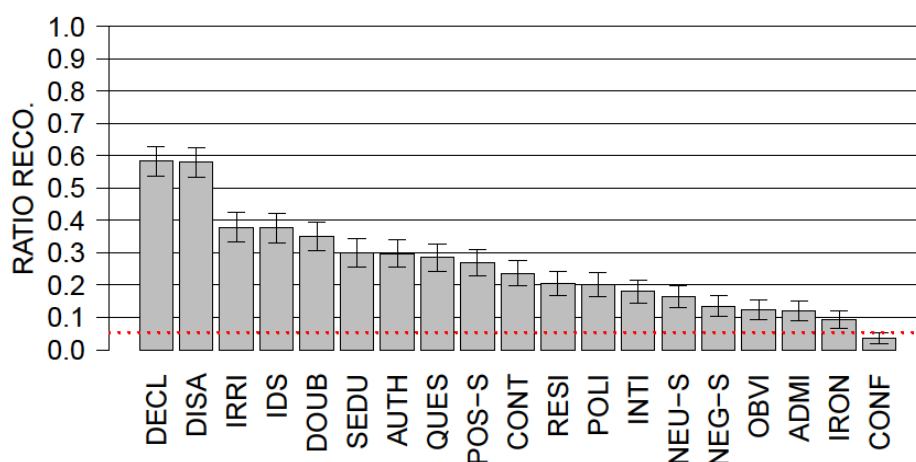


Figure 8 : Taux d'identification des 19 affects sociaux chinois pour 30 sujets natifs. Les affects sociaux sont présentés en abscisse et les scores de reconnaissance en ordonnée. La ligne rouge pointillée indique le seuil du hasard (5,3%).

### 2.3.5.3. Effet de la longueur

Selon le résultat du test post hoc de Tamhane, les énoncés de deux, quatre et neuf syllabes sont significativement mieux reconnus que ceux d'une seule syllabe dont le taux moyen de reconnaissance pour tous les affects sociaux est de 18,4% ( $p < 0,05$ ). Par contre, les scores des stimuli de deux, quatre et neuf syllabes ne montrent pas de différence significative entre eux ( $p > 0,05$ ): 26,8% pour ceux de deux syllabes, 28,0% pour ceux de neuf syllabes et 28,4% pour ceux de quatre syllabes. Cette observation fait écho à l'effet de la longueur des énoncés dans la reconnaissance des affects sociaux dans d'autres langues (Grépillat, 1996; Shochi, 2008; Mac, 2012). Cela pourrait s'expliquer par le fait que les énoncés d'une seule syllabe sont trop courts pour transmettre suffisamment d'informations affectives, notamment en chinois mandarin où le ton est systématique.

Par ailleurs, comme a montré le résultat de l'ANOVA, l'interaction entre l'affect social et la longueur a un effet significatif sur la reconnaissance des affects sociaux, cela signifie que pour les différents affects sociaux, même si leur taux de reconnaissance varie selon la longueur des énoncés, leur façon de varier n'est pas forcément la même (figure 9). Par exemple, bien que le taux de reconnaissance des stimuli les plus courts soit le plus bas, le maternel (IDS) et l'irritation (IRRI) ont suivi une autre loi. Dans le cas de *maternel*, les stimuli d'une et de deux syllabes sont mieux reconnus que ceux de quatre et neuf syllabes (confondus particulièrement avec la *séduction*). Une explication possible

pourrait venir du fait que les adultes adressent rarement des phrases longues et complexes aux jeunes enfants. Pour l'*irritation*, les stimuli de deux syllabes sont moins bien reconnus que les autres et sont confondus avec la *déclaration* et la *confiance*. En écoutant attentivement les stimuli d'*irritation* de toutes les longueurs, il apparaît que la locutrice utilise une qualité de voix rauque et tendue. Par conséquent, les confusions devraient être liées plus aux contours intonatifs qu'à la qualité de voix. Nous allons vérifier cette hypothèse dans la section suivante.

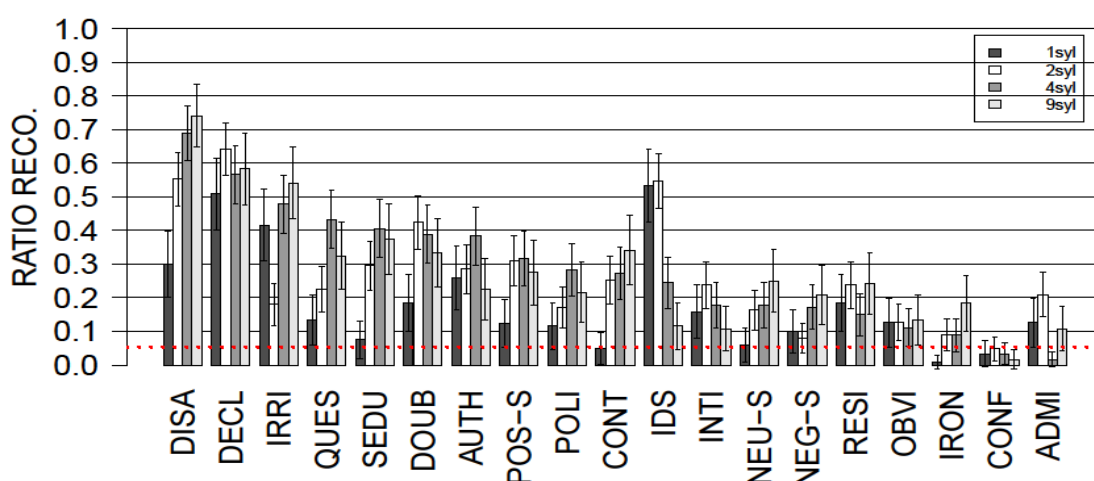


Figure 9 : Effet de la longueur sur les taux d'identification de 19 affects sociaux chinois pour 30 sujets natifs. Les affects sociaux sont présentés en abscisse et les scores de reconnaissance en ordonnée. La ligne rouge pointillée indique le seuil du hasard (5,3%).

#### 2.3.5.4. Effet du genre

Selon le résultat de l'ANOVA, l'effet du genre sur la reconnaissance des affects sociaux est de peu d'ampleur par rapport aux autres facteurs. Cependant, en regardant le taux d'identification des sujets masculins et féminins, représenté par la figure 10, nous remarquons quelques nuances : les femmes (27,2%) reconnaissent légèrement mieux les affects sociaux que les hommes (22,8%), à l'exception de la *déclaration* (DECL), de la *politesse* (POLI) et de l'exclamation de *surprise neutre* (NEU-S).

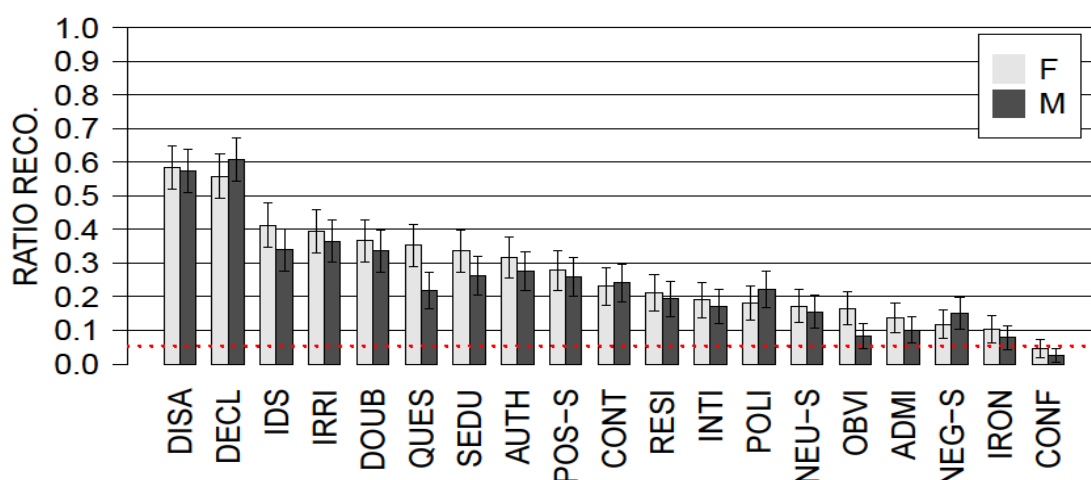


Figure 10 : Effet du genre sur les taux d'identification des 19 affects sociaux chinois pour 30 sujets natifs. Les affects sociaux sont présentés en abscisse et les scores de reconnaissance en ordonnée. Dans la légende, la lettre F indique les femmes et M les hommes. La ligne rouge pointillée indique le seuil du hasard (5,3%).

#### 2.3.5.5. Confusion perçue des affects sociaux

Ici, nous allons plus loin dans l'analyse des confusions perçues entre les affects sociaux, afin de comprendre les difficultés que les sujets éprouvaient durant la perception.

L'analyse de la matrice de confusion montrant les confusions inter-attitudes est résumée à la figure 11. Nous proposons de mettre en évidence les confusions dont le taux est deux fois supérieur au seuil du hasard, - méthode de visualisation proposée pour les affects sociaux de l'anglais par Diaferia (2002). Selon ce mode de seuillage des données, la plupart des reports se font vers la *déclaration* (DECL), notamment la *confiance* (CONF) (56%) et la *politesse* (POLI) (49%) – la *déclaration* elle-même étant bien reconnue (58%), montre quelques confusions avec l'*évidence* (OBVI). La *confiance* et l'*ironie* (IRON) sont mal reconnues. Les trois expressions de *surprise* sont confondues avec le *doute* (DOUB), qui est confondu avec la *question* (QUES) (et vice versa). Le *mépris* (CONT) est confondu avec l'*ironie* et se report aussi vers la *déclaration*. Le *maternel* (IDS) est particulièrement confondu avec la *séduction* (SEDU), ainsi que la *résignation* (RESI) avec la *déception* (DISA).

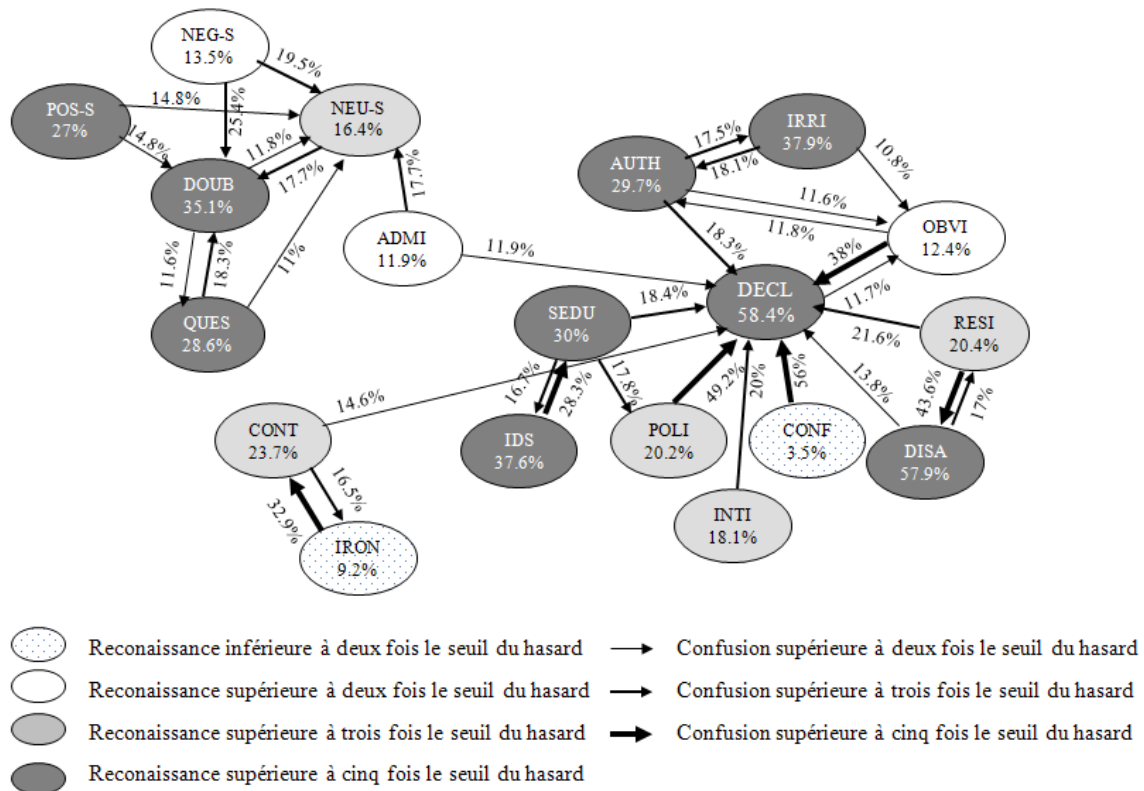


Figure 11 : Confusions perçues des 19 affects sociaux chinois par les sujets natifs : les pourcentages d'identification sont indiqués dans les ovales et les taux de confusions sur les flèches. Seules les confusions qui affichent des scores supérieurs à deux fois le seuil du hasard sont rapportées.

#### 2.3.5.6. Classement des affects sociaux

Afin d'identifier des catégories perceptives plus larges d'affects sociaux pour les deux groupes de sujets (hommes et femmes), un algorithme de classification hiérarchique a été utilisé sur la matrice de confusion. Les distances entre affects sociaux ont été calculées en utilisant la corrélation entre les lignes (la valeur de  $1-r$  est utilisée comme distance, où  $r$  représente la corrélation entre deux lignes de la matrice de confusion). Le classement des 19 affects sociaux est illustré à la figure 12.

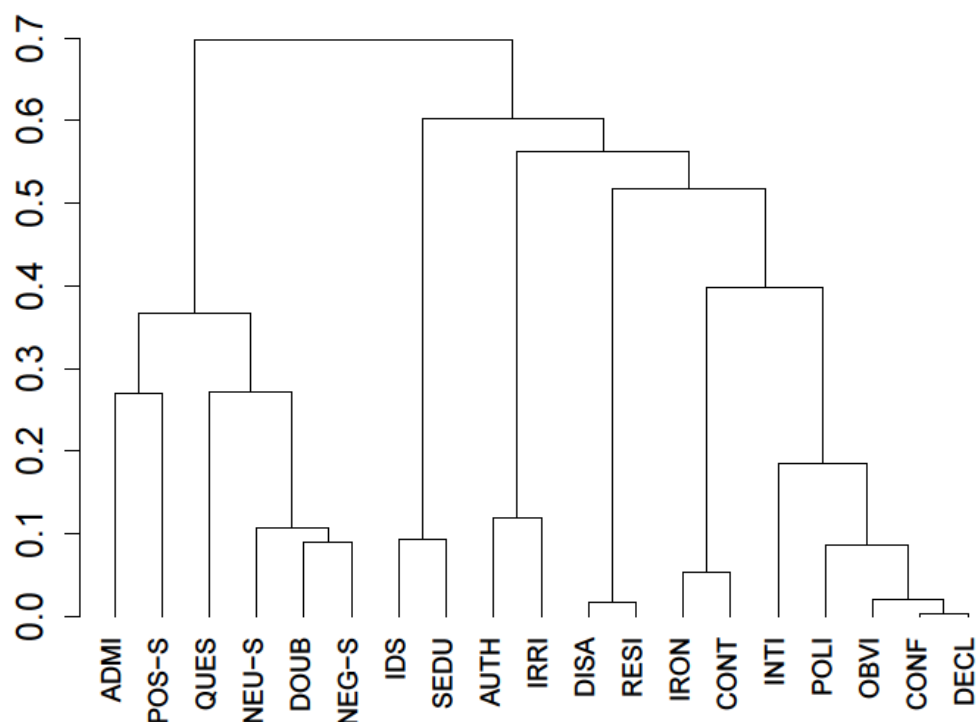


Figure 12 : Classification hiérarchique des 19 affects sociaux chinois pour les sujets natifs.

Au plus haut niveau du dendrogramme, les affects sociaux sont divisés en deux groupes principaux : un des groupes rassemble la *déclaration*, la *confiance*, l'*évidence*, la *politesse*, l'*intimité*, le *mépris*, l'*ironie*, la *résignation*, la *déception*, l'*irritation*, l'*autorité*, la *séduction* et le *maternel*, alors que l'autre groupe est plutôt composé des trois expressions de surprise, du *doute*, de la *question* et de l'*admiration*. Cette séparation correspond bien au résultat de l'analyse brute sur les taux de confusions des affects sociaux (voir figure 11).

Dans le cadre d'une description plus fine, les 19 affects sociaux sont répartis en sept groupes :

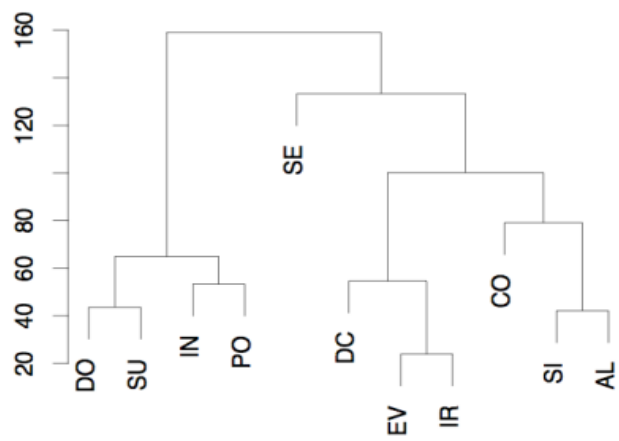
- L'*admiration* et l'exclamation de *surprise positive* : ces affects sont caractérisés par une évaluation de valence très positive pour un fait ou une personne.
- L'*exclamation de surprise négative*, de *surprise neutre*, la *question* et le *doute* : ces affects ont en commun une évaluation mentale modifiée (l'exclamation de *surprise*) ou modifiable (le *doute* et la *question*) d'un fait, ce qui pourrait être analogue à l'inconfort dans la croyance du fait énoncé. Ils peuvent aussi être considérés comme des expressions inattendues (l'exclamation de *surprise*



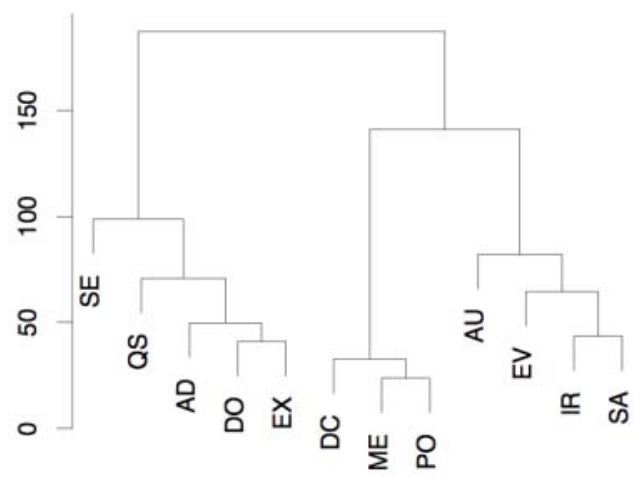
*négligative* et de *surprise neutre*), interrogatives (la *question*) et incertaines (le *doute*).

- Le *maternel* et la *séduction* : ce sont des expressions d'un statut durable de proximité (le *maternel*) ou d'un souhait de proximité (la *séduction*). Les deux expressions utilisent une voix soufflée avec des implications de soin et d'intimité (Wichmann, 2000; Campbell, 2004).
- L'*autorité* et l'*irritation* : deux actes de langage imposant la volonté du locuteur à son interlocuteur.
- La *déception* et la *résignation* : en opposition à l'*autorité* et l'*irritation*, elles sont des attitudes négatives, soumises et conceptuellement proches.
- L'*ironie* et *mépris*, qui pourraient être décrites comme des expressions de refus avec lesquelles le locuteur exprime son désaccord avec son interlocuteur.
- La *déclaration*, la *confiance*, l'*évidence*, la *politesse* et l'*intimité* : elles constituent une catégorie de différentes manières d'affirmer quelque chose. La *politesse* et l'*intimité* se distinguent des trois autres attitudes d'assertion par leur usage relativement social et situationnel.

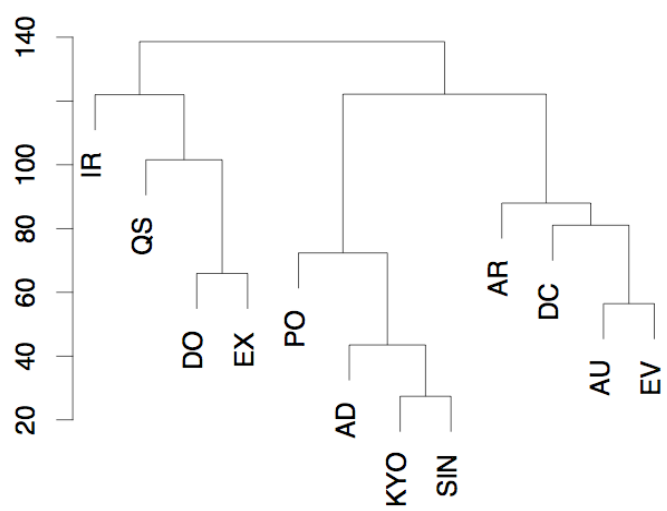
En comparant la classification des affects sociaux chinois avec celle des affects sociaux en anglais, français, japonais et vietnamien (figure 13), nous constatons que les affects sociaux de langues différentes sont classés de manière similaire, et que plus de similarités sont observées entre la classification des affects sociaux chinois et vietnamiens. La *surprise*, le *doute* et la *question* sont acoustiquement jugés proches par les auditeurs de ces cinq langues sauf le vietnamien ; la *déclaration* est généralement confondue avec la *politesse* et l'*autorité* avec l'*irritation* par les auditeurs chinois, français et vietnamiens; par contre, l'*ironie* et le *mépris* sont confondus par les auditeurs chinois, vietnamiens et anglais.



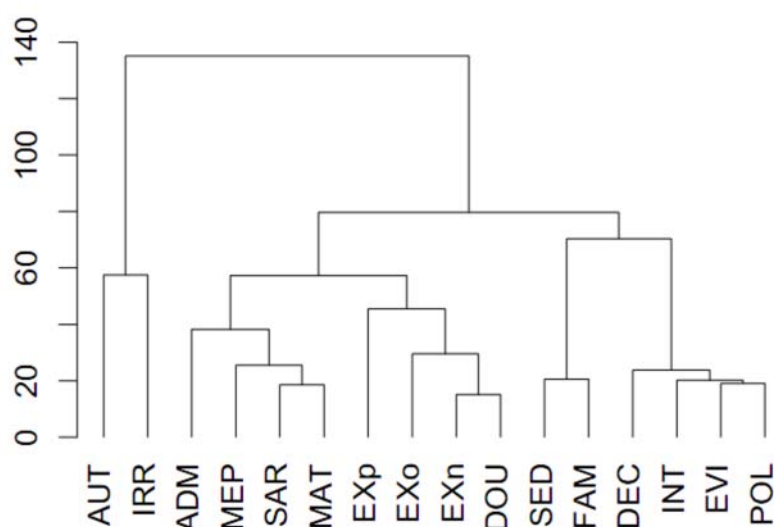
(a)



(b)



(c)



(d)

Figure 13 : Classification des affects sociaux en anglais (a), français (b), japonais (c) et vietnamien (d).

Signification des étiquettes : AD (admiration), AU-AUT-CO (autorité), DC-DEC (déclaration), DO-DOU (doute), EV-EVI (évidence), EX-SU (exclamation de surprise), EXp (exclamation de surprise positive), Exo (exclamation de surprise neutre), EXn (exclamation de surprise négative), IN-FAM (intimité-familiarité), IR-IRR (irritation), KYO (kyoshuku), ME-MEP-AL-AR (mépris-arrogance), MAT(maternel), PO-POL (politesse), QS (question), SE (séduction), SI-SA-SAR (ironie sarcastique), SIN (sincérité) (Shochi, 2008; Mac, 2012).

### 2.3.6. Conclusion

Les 21 énoncés exprimés avec 19 affects sociaux (399 stimuli) ont été sélectionnés pour la validation d'un corpus des affects sociaux chinois composé de 2888 stimuli. La plupart des affects sociaux sont reconnus par les 30 auditeurs chinois natifs, sauf l'attitude de *confiance* dont le taux d'identification est inférieur au seuil du hasard. Les taux d'identification moyens des affects sociaux ne sont pas très élevés et cela serait dû à la complexité cognitive de la tâche de reconnaissance d'une étiquette parmi 19, mais aussi par un effet possible du contenu sémantique des phrases. Plusieurs sujets ont signalé leur sentiment d'un décalage entre le contenu sémantique (neutre) de l'énoncé et les affects sociaux exprimés. Par ailleurs, une tâche de reconnaissance prosodique hors contexte est toujours délicate et entachée de bruit.

La longueur des énoncés et l'interaction entre la longueur et l'affect social montrent un effet significatif ( $p < 0,01$ ) sur la reconnaissance des affects sociaux. Précisément, la différence de taux d'identification entre les énoncés d'une syllabe et ceux des autres

longueurs est significative ( $p < 0,01$ ). Cependant, la différence entre les longueurs de deux, quatre et neuf syllabes ne l'est pas. Le taux moyen d'identification des énoncés d'une seule syllabe est plus bas que celui des autres et les énoncés de quatre syllabes ont reçu le meilleur score moyen de reconnaissance. Le genre induit aussi un effet significatif ( $p < 0,01$ ) mais de peu d'ampleur par rapport aux autres facteurs. Globalement, les femmes (27,2%) reconnaissent mieux les affects sociaux que les hommes (22,8%), à l'exception de la *déclaration*, la *politesse* et l'exclamation de *surprise neutre*.

L'observation des confusions perceptives montre que la *déclaration* attire la plupart des confusions, ce qui correspond aux résultats de Diaferia (2002) et de Mac et al. (2010). Vu la complexité de la tâche, le choix de la *déclaration* peut constituer la catégorie refuge la plus neutre pour les sujets. A la place des seuls indices acoustiques, des indices visuels peuvent être très utiles pour de pareilles identifications (Rilliard et al., 2009; de Moraes & Rilliard, 2014). Basée sur la matrice de confusion, la classification hiérarchique des résultats de perception regroupe les affects sociaux en sept catégories, et au sein de chacune, les affects sociaux sont cohérents au niveau de processus cognitif.

## **2.4. Analyse acoustique des affects sociaux chinois**

Après la validation du corpus de 19 affects sociaux chinois, nous poursuivons, dans cette section, à analyser les principales caractéristiques prosodiques de ces expressions socio-affectives à la fois au niveau global des paramètres prosodiques classiques et de la forme du contour de F0 de l'énoncé.

### **2.4.1. Introduction**

Il ne fait pas de doute que les signaux acoustiques véhiculent les expressions émotionnelles et attitudinales du locuteur. La mesure des paramètres acoustiques permet d'observer comment ces signaux affectifs modifient la voix humaine (Deller et al., 2000; Yuan et al., 2002; Zhang et al., 2006; Raphael et al., 2007; Scherer & Ellgring, 2007). Dans cette section, nous essayons dans un premier temps d'identifier les principales caractéristiques prosodiques des 19 affects sociaux du chinois mandarin en mesurant leur proximité en fonction des valeurs des paramètres acoustiques de fréquence fondamentale, de durée et d'intensité. Ensuite, une autre analyse du contour intonatif des affects sociaux est mise en place pour les mêmes signaux acoustiques. Ces

analyses consistent également à examiner le lien entre les traits prosodiques et les catégories perceptives des affects sociaux identifiées dans l'expérience perceptive (voir supra).

#### 2.4.2. Sous-corpus

76 stimuli, à savoir quatre énoncés exprimés avec 19 affects sociaux, qui ont été validés dans l'expérience perceptive précédente, sont retenus pour cette analyse acoustique. Ces énoncés ont reçu le meilleur taux moyen de reconnaissance pour tous les affects sociaux et pour chaque longueur (monosyllabe, dissyllabe, quadrisyllabe et neuf syllabes). Les stimuli représentent donc les meilleures expressions prototypiques de ces affects sociaux. La table 7 présente les énoncés sélectionnés pour le sous-corpus. Les chiffres dans la colonne Tons représentent la combinaison des tons de l'énoncé : 1 fait référence à un ton plat, 2 à un ton montant, 3 à un ton descendant-remontant, 4 à un ton descendant et 0 à un ton neutre.

Nombre de syllabes	Tons	Phrases en chinois	Transcription en pinyin	Traduction française
1	4	树	shu4	arbre
2	4-2	放学	fang4 xue2	l'école est finie
4	1-1-1-2	张医生来	Zhang1 yi1 sheng1 lai2	Docteur Zhang viendra
9	2-1-1-1-1-1-1-1-4	王医生他三姑妈休假	Wang2 yi1 sheng1 ta1 san1 gu1 ma1 xiu1 jia4	La troisième tante paternelle de Docteur Wang prend un congé.

Table 7: Énoncés sélectionnés pour l'analyse acoustique des 19 affects sociaux chinois.

#### 2.4.3. Paramètres prosodiques et extraction des valeurs acoustiques

Puisque la fonction attitudinale de la prosodie couvre la taille de l'énoncé, la prosodie des affects sociaux pourrait se concrétiser dans le contour global de l'énoncé (Aubergé, 2002a). Cette proposition a été confirmée par les études de Li et al. (2011) et de Lin & Fon (2012), dans lesquelles la forme globale de l'intonation observée est étroitement liée à sa fonction expressive. Mac (2012) a montré, dans ses travaux de thèse, que par rapport au mouvement du contour de la fréquence fondamentale au milieu de l'énoncé, la variation des mouvements initiaux et finaux sont plus significatifs pour caractériser le contour de fréquence fondamentale des différentes attitudes vietnamiennes.

En référence aux travaux cités ci-dessus, nous procédons, lors de cette analyse, aux mesures prosodiques des 19 affects sociaux chinois à la fois sur l'ensemble de l'énoncé et sur la première et la dernière syllabe.

Parmi les paramètres acoustiques étudiés dans la littérature, par exemple la fréquence fondamentale, l'intensité, la durée, la structure des harmoniques, la pente spectrale, la fréquence fondamentale (F0), l'intensité et la durée sont les paramètres les plus souvent utilisés comme corrélats de la prosodie (Scherer & Ellgring, 2007). Notamment, la fréquence fondamentale est consensuellement considérée comme un indice important, qui signale l'attitude, l'intention ou l'émotion du locuteur (Ohala, 1994). La durée, qui constitue le support physique temporel du signal de la parole, est une composante essentielle de la prosodie (Di Cristo, 2013, p. 89). Au contraire à la fréquence fondamentale et à la durée, l'intensité est un paramètre qui ne fait pas l'objet de soins attentifs dans la littérature prosodique pour une raison éventuelle de la difficulté à mesurer et à son évaluation subjective (Di Cristo, 2013, p. 89). Partant de ces réflexions, nous avons extrait pour les 76 stimuli considérés des valeurs de ces trois paramètres parmi lesquels la fréquence fondamentale et la durée sont au centre de nos observations.

La plupart des paramètres acoustiques de la prosodie sont extraits automatiquement à partir des signaux à l'aide du logiciel PRAAT (Boersma, 2001). Chaque signal est manuellement étiqueté au niveau phonémique, et les contours de F0 sont normalisés pour leur variation temporelle en extrayant dix points pour chaque voyelle. Pour chaque énoncé, un ensemble d'indices prosodiques est calculé sur la base de ces trois paramètres acoustiques. Pour la fréquence fondamentale, sont considérés sa valeur moyenne, son écart-type, les valeurs maximum et minimum, les valeurs moyennes de la première et de la dernière syllabe. Pour la durée, sont calculées la durée de la première et de la dernière syllabe, et celle de l'énoncé complet. Enfin la valeur moyenne de l'intensité de l'énoncé est aussi prise en compte.

Les valeurs de F0 sont estimées par l'algorithme de cross-corrélation du logiciel PRAAT et sont exprimées en demi-tons en prenant 1 Hz comme valeur de référence. L'intensité est exprimée en décibels (dB). Le logarithme décimal de la durée exprimée en milliseconde est calculé et utilisé dans l'analyse (Ross et al., 1986). Deux indices supplémentaires sont également calculés :

- la plage des valeurs de F0 (exprimée en demi-tons), définie comme l'écart entre la valeur maximum et minimum de F0 pour un énoncé considéré (Lin & Fon, 2012);
- la pente de F0 (exprimée en demi-tons/s) et définie comme la différence entre la valeur moyenne de la dernière syllabe et celle de la première syllabe de l'énoncé, divisée par la durée de la phrase (Ross et al., 1986).

Les valeurs moyennes des paramètres acoustiques sélectionnés pour les expressions vocales des 19 affects sociaux chinois sont calculées avant d'être statistiquement analysées. La table 8 résume les dix variables prosodiques extraites des échantillons sonores.

Indice prosodique	Unité	Abréviation	Paramètre acoustique
plage de F0	demi-tons	F0_range	F0
moyenne de F <sub>0</sub> de la phrase	demi-tons	F0_mean	F0
écart-type de F0	demi-tons	F0_std	F0
moyenne de F <sub>0</sub> de la première syllabe	demi-tons	F_F0_mean	F0
moyenne de F <sub>0</sub> de la dernière syllabe	demi-tons	L_F0_mean	F0
pente de F0	demi-tons /s	F0_slope	F0
moyenne d'intensité de l'énoncé	dB	Intensity_mean	intensité
durée de la première syllabe	log <sub>10</sub> (ms)	F_dur	durée
durée de la dernière syllabe	log <sub>10</sub> (ms)	L_dur	durée
durée de l'énoncé	log <sub>10</sub> (ms)	Sentence_dur	durée

Table 8 : Indices prosodiques calculés sur la base des paramètres acoustiques de fréquence fondamentale, d'intensité et de durée.

#### 2.4.4. Analyse statistique et résultats

##### 2.4.4.1. Classement des affects sociaux

En utilisant l'ensemble de valeurs des indices prosodiques décrits ci-dessus, nous avons mené des analyses statistiques dans le but de mesurer, d'un côté, la différence entre les affects sociaux en fonction des valeurs de variables acoustiques, et de l'autre côté, l'apport de ces différents indices à la perception des affects sociaux. Deux analyses ont été réalisées à l'aide du logiciel SPSS 18.0 : une Analyse en Composantes Principales

(ACP), qui permet de visualiser l'importance relative de différents indices prosodiques sur la réalisation de chaque affect social ; l'autre est une classification hiérarchique pour mesurer les distances acoustiques entre les 19 expressions socio-affectives et vérifier si ces distances recoupent les distances perceptives. Nous décidons de combiner les deux analyses en effectuant une classification hiérarchique sur les composantes principales obtenues en sortie d'une analyse factorielle (HCPC), parce que cette méthode a l'avantage de regrouper les individus avec moins de bruit (Husson et al., 2010). Pour ce faire, en tant qu'une étape de prétraitement avant l'analyse de classification, une ACP est menée sur les données en croisant les dix indices prosodiques aux 19 affects sociaux.

Nous mesurons préalablement l'adéquation de l'échantillonnage (*sampling adequacy* en anglais) à une ACP à l'aide d'un test de Bartlett et de Kaiser-Meyer-Olkin (*KMO & Bartlett's test* en anglais). Le résultat montre que chacune des variables considérées a un certain degré de corrélation avec les autres ( $KMO = 0,614$ ;  $Bartlett's\ Sig. < 0,01$ ). Ceci justifie la pertinence des données à une ACP. Le résultat du test de Bartlett et de KMO est rapporté dans la table 7.

Kaiser-Meyer-Olkin Measure of Sampling Adequacy		0,614
Bartlett's Test of Sphericity	Approx. Chi-Square	263,109
	dll	45
	Sig.	0,000

Table 9 : Résultat du test de Bartlett et de Kaiser-Meyer-Olkin montrant l'adéquation des données à une ACP.

Ensuite, une ACP est effectuée sur les valeurs standardisées. D'après le résultat, les deux composantes principales expliquent 82% de la variance (table 10). Pour examiner les corrélations entre les deux facteurs principaux et les variables acoustiques, la méthode de rotation varimax est utilisée. Ce qui nous permet d'observer la différenciation des deux facteurs, et ainsi de les interpréter de façon pertinente. Les corrélations factorielles après rotation, qui sont présentées dans la table 11, indiquent que le premier facteur principal est essentiellement lié aux indices de F0 et à l'intensité moyenne, tandis que le second facteur porte plutôt sur les indices de durée et sur la pente de F0 (et donc sa dynamique).



Component	Initial Eigenvalues		
	Total	% of Variance	Cumulative %
1	5.900	59.003	59.003
2	2.295	22.954	81.957
3	.903	9.027	90.984
4	.523	5.227	96.211
5	.163	1.628	97.839
6	.112	1.123	98.962
7	.066	.655	99.617
8	.027	.269	99.886
9	.008	.081	99.967
10	.003	.033	100.000

Table 10 : Proportion de la variance expliquée par chaque composante de l'ACP sur les 10 indices prosodiques.

	Component 1	Component 2
$F_0$ _range	0.95	0.17
$F_0$ _mean	0.93	0.05
$F_0$ _std	0.83	0.20
F_ $F_0$ _mean	0.97	0.07
L_ $F_0$ _mean	0.95	0.07
F_dur	0.20	0.90
L_dur	0.70	0.64
Sentence_dur	0.28	0.94
$F_0$ _slope	-0.42	0.71
Intensity_mean	0.67	0.01

Table 11 : Matrice des scores factoriels des 10 indices prosodiques après la rotation varimax.

Les scores factoriels des 19 affects sociaux sont sauvegardés et puis utilisés comme une nouvelle variable sur laquelle une analyse de classification hiérarchique agglomérative a été lancée. Le carré de la distance Euclidienne entre les observations a été utilisé pour regarder les classes différentes, en utilisant le critère de variance minimale de Wald. En vue de déterminer le nombre optimal de groupes sur l'arbre hiérarchique, nous nous sommes référés au critère de réduction de variance (*elbow criterion* en anglais) expliquée par chaque groupe (Soni Madhulatha, 2012). La figure 14 expose le dendrogramme résultant de la classification hiérarchique et le graphique de perte d'inertie en fonction

du nombre de groupes (en haut à droite), selon lequel il est raisonnable de considérer 5 groupes d'attitudes.

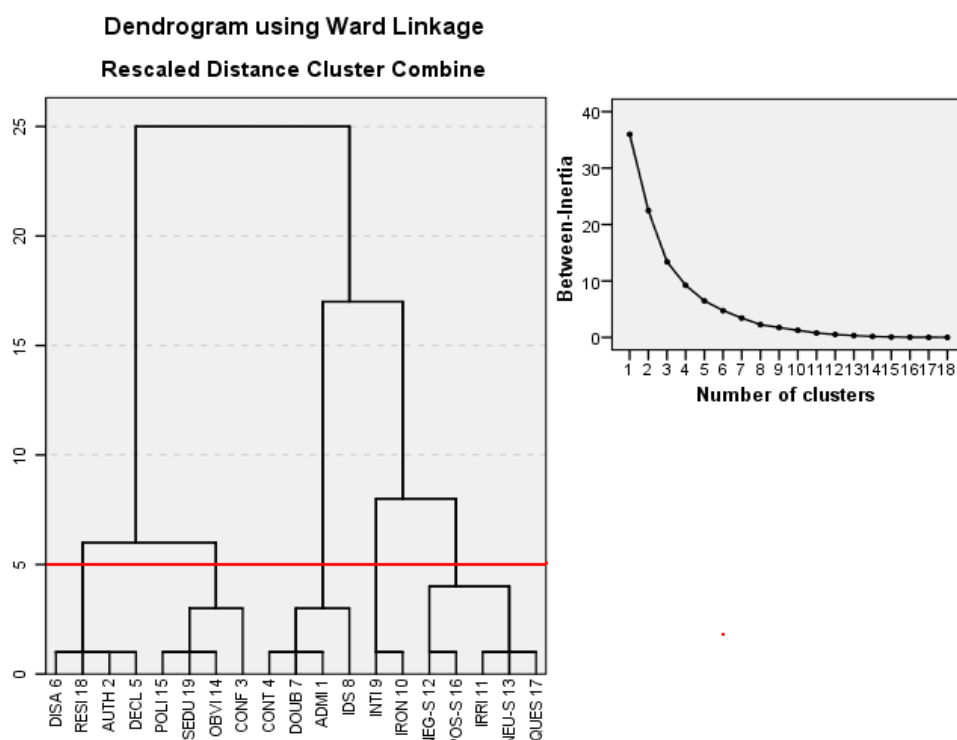


Figure 14 : Classification hiérarchique des 19 affects sociaux à partir de leurs indices prosodiques. La ligne rouge indique les groupes obtenus selon la réduction de variance présentée dans le graphique en haut à droite.

La dernière étape de cette analyse HCPC consiste à positionner les affects sociaux regroupés en cinq groupes sur l'espace composé des deux facteurs principaux. Ce qui a pour objectif de montrer comment les affects sociaux se dispersent sur ces deux dimensions, ainsi que d'observer l'apport des indices prosodiques sur le regroupement des affects sociaux.

Comme illustré sur la figure 15, les affects sociaux se dispersent le long des axes de coordonnées ( $x=0$ ,  $y=0$ ) et certains sont placés à l'extrémité de ces deux axes : sur la première dimension, l'exclamation de *surprise positive* (POS-S), de *surprise négative* (NEG-S) et de *surprise neutre* (NEU-S) se regroupent vers le score factoriel le plus élevé (2,00 sur le graphique), tandis que les attitudes de *déception* (DISA), de *résignation* (RESI) et de *confiance* (CONF) montrent les scores les plus faibles (scores négatifs); sur la seconde dimension, la valeur du *maternel* (IDS) est la plus élevée contrairement à

celle de *l'intimité* (INTI). Par ailleurs, certaines paires d'affects sociaux partagent des coordonnées relativement similaires dans cet espace graphique. Il s'agit du couple de l'exclamation de *surprise positive* (POS-S) et de *surprise négative* (NEG-S), de la *déception* (DISA) et la *résignation* (RESI), de la *séduction* (SEDU) et la *politesse* (POLI), de *l'autorité* (AUTH) et la *déclaration* (DECL). Cette proximité entre deux affects sociaux impliquant la similarité entre leurs caractéristiques prosodiques, pourrait expliquer les confusions observées en perception (voir supra).

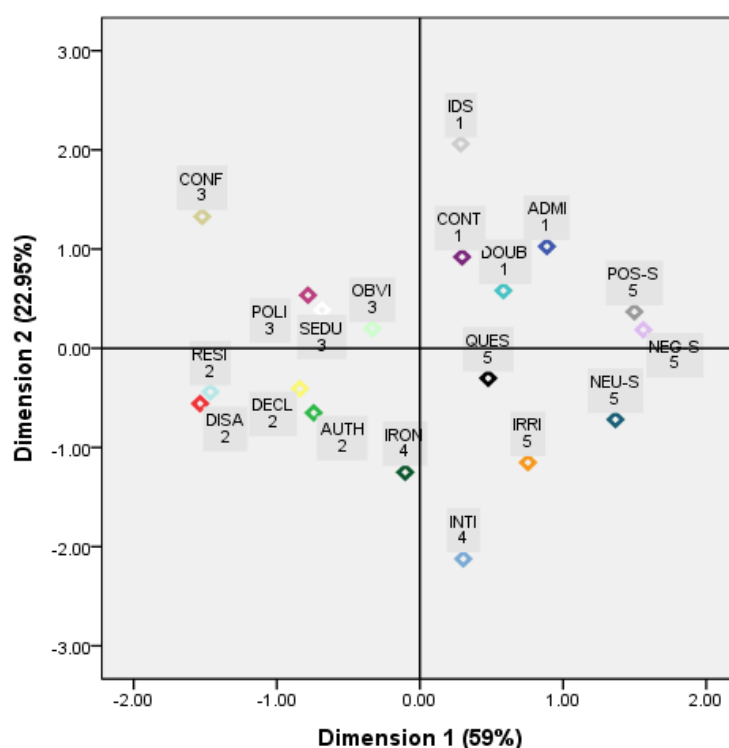


Figure 15 : Représentation des deux dimensions principales de l'ACP, qui expliquent 82% de la variance et sur lesquelles les affects sociaux sont projetés. Les chiffres en-dessous des étiquettes indiquent à quel groupe cet affect social appartient.

En ce qui concerne les groupes d'affects sociaux obtenus grâce à la classification hiérarchique, nous constatons que les 19 affects sociaux se répartissent en 5 groupes sur les deux dimensions principales :

- Le groupe 1 regroupe des affects sociaux d'*admiration* (ADMI), de *maternel* (IDS), de *mépris* (CONT) et de *doute* (DOUB), qui montrent des valeurs élevées sur les paramètres de F0 et de durée.

- Le groupe 2 réunit l'*autorité* (AUTH), la *déclaration* (DECL), la *résignation* (RESI) et la *déception* (DISA), qui, au contraire du premier groupe, présentent des valeurs faibles sur les paramètres de F0 et de durée.
- Le groupe 3, composé de l'*évidence* (OBVI), de la *séduction* (SEDU), de la *politesse* (POLI) et de la *confiance* (CONF), manifeste des similarités avec les affects sociaux du deuxième groupe sur le plan de fréquence fondamentale, mais se différencie avec des valeurs plus élevées en durée.
- Le groupe 4 comprend l'*ironie* (IRON) et l'*intimité* (INTI), qui sont assez isolées des autres affects sociaux et s'identifient par leurs valeurs extrêmement faibles de durée.
- Le groupe 5 inclut les trois exclamations de *surprise*, la *question* (QUES) et l'*irritation* (IRRI). Ce groupe manifeste des similarités avec le premier groupe sur le plan de la F0, et parallèlement des différences sur le plan de la durée.

#### 2.4.4.2. Caractéristiques acoustiques des affects sociaux

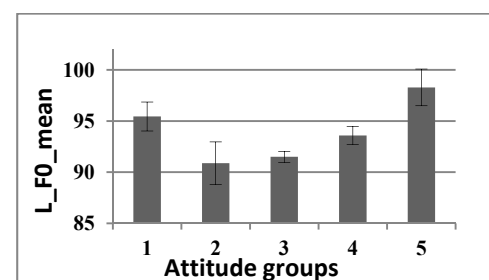
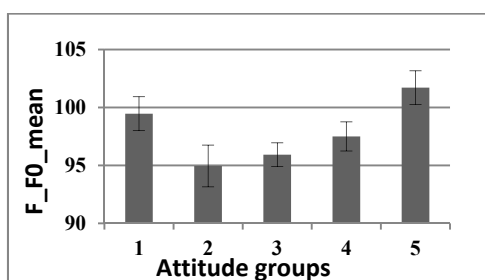
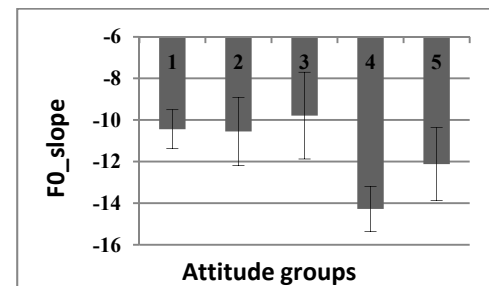
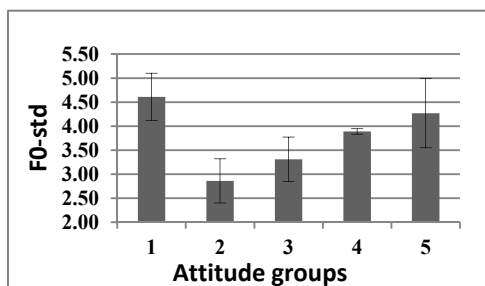
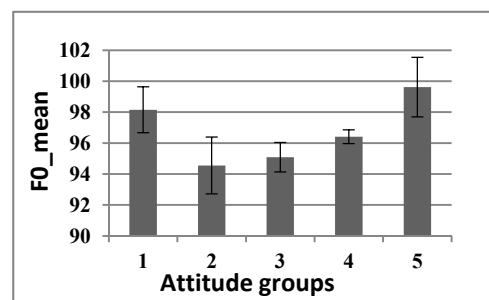
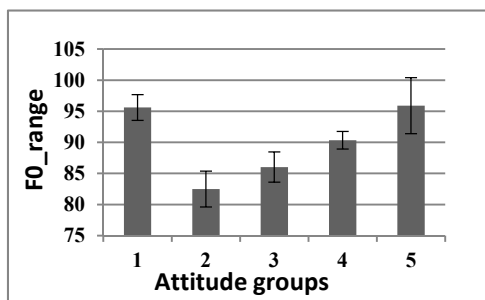
La classification hiérarchique effectuée sur les composantes principales nous a donné une image globale sur la dispersion des affects sociaux et de leur regroupement en fonction des indices prosodiques. Afin d'observer en détails les traits prosodiques de chaque groupe, autrement dit, les différences prosodiques entre les groupes, nous recourrons à une comparaison de la valeur moyenne de chaque variable prosodique pour chaque groupe. Le résultat est détaillé par les graphiques de la figure 16.

En ce qui concerne les indices de F0 (à savoir la plage de F0, la F0 moyenne de la phrase, la F0 moyenne de la première et dernière syllabe, l'écart-type des valeurs de F0 de l'énoncé et la pente de F0), les groupes 1 et 5 possèdent des valeurs plus élevées que les autres groupes, et les groupes 2 et 3 des valeurs plus petites. Précisément, le groupe 5 montrent les valeurs les plus élevées en termes de plage de F0, de registre de F0 de l'énoncé et de la première et dernière syllabe ; le groupe 1 a la valeur la plus élevée en variation de F0 de l'énoncé ; le groupe 2 a les valeurs les plus petites pour presque tous les indices de F0 sauf la pente de F0 pour laquelle le groupe 4 montre la valeur négative la plus forte. Tous les groupes se montrent légèrement différents sur le plan du registre de F0 au niveau de l'énoncé. Par ailleurs, pour différencier les groupes qui se ressemblent dans le domaine de F0, nous nous référons à l'aspect de timing : pour les

affects sociaux du groupe 1, leurs valeurs de durée de la première syllabe et de l'énoncé sont plus élevées que celles observées pour le groupe 5 ; les affects sociaux du groupe 3 ont des valeurs supérieures à celles du groupe 2 pour toutes les mesures de durée mais notamment la durée de la dernière syllabe.

Concernant les indices de durée (à savoir la durée de la première et de la dernière syllabe et la durée de l'énoncé), le groupe 1 montre les valeurs les plus élevées, et les groupes 2 et 4, en revanche, les plus petites. Les groupes 3 et 5 se situent entre les deux, tandis que le groupe 5 présente la deuxième plus grande valeur en durée de la dernière syllabe après le groupe 1.

Quant à l'intensité (à savoir le registre de l'intensité de l'énoncé), tous les groupes se montrent assez homogènes sans faire ressortir de tendance claire.



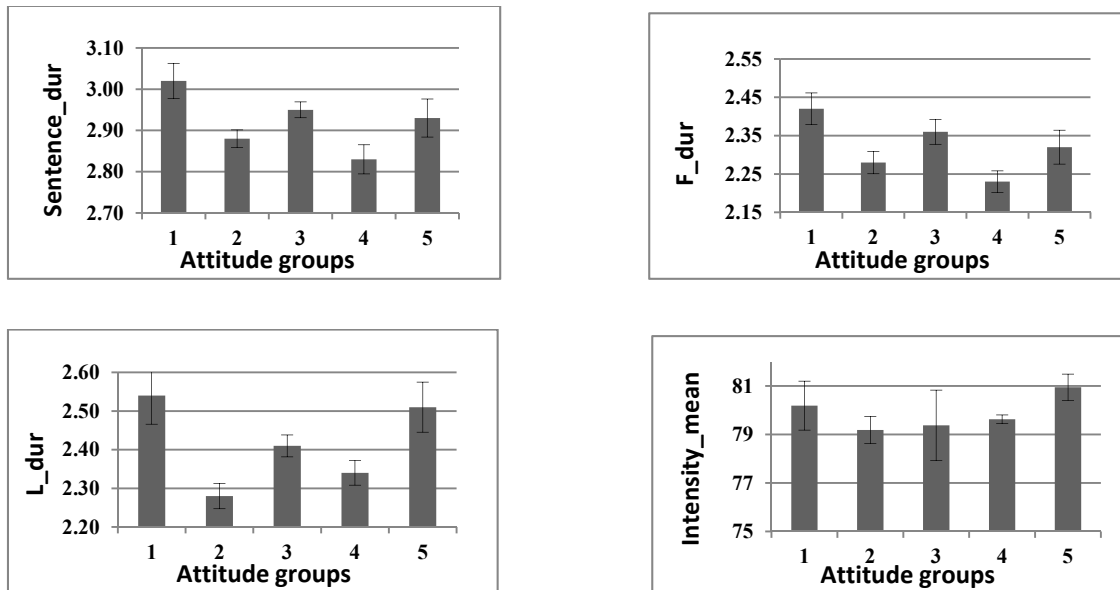


Figure 16 : Valeurs moyennes des indices prosodiques pour les 5 groupes d'affects sociaux. Les chiffres sur l'abscisse représentent les groupes obtenus dans la classification hiérarchique, et ceux au-dessus des barres les valeurs moyennes d'un paramètre acoustique donné.

Pour conclure, à partir de l'analyse présentée ci-dessus, nous essayons dans ce paragraphe de résumer les caractéristiques prosodiques saillantes de chaque groupe d'affects sociaux particulièrement en fonction des indices de F0 et de durée : le groupe 1 (l'*admiration*, le *maternel*, le *mépris* et le *doute*) est caractérisé par un allongement marqué sur la première et la dernière syllabe comme sur l'ensemble de l'énoncé, ainsi que par de hautes valeurs des indices de F0, tandis que le groupe 2 (l'*autorité*, la *déclaration*, la *résignation* et la *déception*) se caractérise par des valeurs à l'opposé (durées courtes et F0 basse); le groupe 5 (l'exclamation de *surprise positive*, de *surprise négative*, de *surprise neutre*, la *question* et l'*irritation*) est typiquement marqué par une large plage de F0 et un niveau élevé de F0 moyenne aussi bien sur la première et la dernière syllabe que sur l'ensemble de l'énoncé, ainsi que par une durée moyenne importante sur la dernière syllabe; le groupe 4 (l'*ironie* et l'*intimité*) s'identifie par ses plus petites valeurs de durée et donc par une énonciation rapide et une importante pente descendante de F0 au long de l'énoncé ; les affects sociaux du groupe 3 (l'*évidence*, la *séduction*, la *politesse* et la *confiance*) montrent des similarités avec ceux du groupe 2 en terme de F0, mais ils se distinguent par une énonciation plus lente.

Par ailleurs, à l'intérieur de ces groupes, certains affects sociaux montrent des particularités. Il s'agit du *maternel* dont le débit de parole est le plus lent (Fernald &

Mazzie, 1991), de l'*intimité* dont la durée est la plus courte, des trois expressions de *surprise* qui se distinguent par leurs très hautes valeurs de F0 (Caelen-Haumont, 1991; Kehrein, 2002), et de la *résignation* et la *déception* dont les valeurs en F0 sont les plus faibles en comparaison des autres affects sociaux. La *confiance* s'éloigne des autres membres du groupe 3 en raison de ses valeurs de F0 plus faibles et de sa durée plus longue

#### 2.4.5. Analyse des contours intonatifs des affects sociaux

Lors de l'analyse précédente, nous avons examiné les traits saillants de la prosodie des affects sociaux en les regroupant en catégories différentes. L'avantage de cette méthode est de pouvoir observer la tendance globale de la variation prosodique des affects sociaux, mais son principal inconvénient repose sur l'incapacité d'extraire les caractéristiques prosodiques spécifiques à chaque affect social, pour chaque longueur d'énoncé. En tenant compte de ce fait, nous avons effectué une analyse des contours intonatifs de ces énoncés. L'objectif de cette analyse est, d'une part, d'observer le mouvement de la courbe de F0 de l'expression socio-affective et ainsi de les comparer par paire ; et d'autre part, de tenter d'expliquer les confusions perceptives entre les affects sociaux apparues lors de l'expérience auditive réalisée précédemment.

##### 2.4.5.1. Méthodologie

Pour réaliser ces objectifs, deux mesures de similarité des contours intonatifs ont été utilisées : l'une consiste en une mesure de la similarité de la forme de courbes intonatives des affects sociaux ; l'autre porte sur la distance des registres de F0 entre deux courbes intonatives. Les séquences de valeurs de F0 des contours normalisés à 10 points par voyelle servent de base à ces mesures.

La première mesure a été effectuée par le calcul du coefficient de corrélation pondéré ( $r$ ) entre deux courbes intonatives. La pondération est réalisée en fonction de l'énergie du signal à l'endroit de la mesure. Cette mesure a été proposée par (Hermes, 1998) pour mesurer la similarité et la différence entre les configurations intonatives dans l'évolution temporelle des courbes et est jugée la plus pertinente par rapport à la perception. Rilliard & Lai (2008), Romano & Miotti (2008) et Moutinho et al. (2011) ont adapté cette méthode dans leurs recherches sur la variation prosodique dialectale dans les langues romanes, ainsi que d'Alessandro et al. (2011) qui l'ont appliquée à la

stylisation et la modalisation de l'intonation pour une étude de l'analogie entre le geste de la main et les mouvements prosodiques. Ces études ont confirmé la pertinence de cette méthode pour évaluer la proximité perceptive de deux contours prosodiques, car elle permet de calculer une « distance »<sup>15</sup> entre les formes de différents contours en accordant une plus grande importance aux parties montrant une grande énergie de voisement, qui sont donc plus pertinentes perceptivement. Par conséquent, c'est une mesure dédiée à la comparaison objective de contours intonatifs et à l'évaluation objective de la similarité perceptive entre deux courbes comparables, et elle nous paraît en cela adaptée à notre situation.

Le calcul du coefficient de corrélation pondéré ( $r$ ) a été réalisé à l'aide de la formule proposée par d'Alessandro et al. (2011), dérivée des travaux de Hermes (1998) :

$$r_{f_1 f_2} = \frac{\sum_i w(i)(f_1(i) - m_1)(f_2(i) - m_2)}{\sqrt{\sum_i w(i)(f_1(i) - m_1)^2 \sum_i w(i)(f_2(i) - m_2)^2}}$$

Dans cette formule,  $f_1$  et  $f_2$  représentent les valeurs de F0 (exprimées en demi-tons, référence à 1Hz) des deux contours intonatifs, extraites en dix points de chaque voyelle ;  $m_1$  et  $m_2$  sont les valeurs moyennes de F0 de ces deux phrases ;  $w$  est la pondération en fonction de l'énergie du signal, calculée comme la moyenne des deux valeurs d'intensité mesurées à un point donné pour les deux contours comparés (exprimée en dB) ; l'indice  $i$  représente le nombre de points de F0 pour la phrase entière. Le calcul du coefficient de corrélation pondéré est effectué automatiquement à l'aide de scripts Matlab. Le coefficient varie entre -1 et 1 : lorsque les formes des deux courbes sont plus proches, la corrélation tend vers 1 ; tandis qu'elles sont systématiquement opposées, la corrélation tend vers -1. Les corrélations obtenues dans la mesure sont ensuite transformées à l'aide de la transformée en Z de Fisher (voir Hermes, 1998 pour des détails).

L'autre mesure porte sur le calcul de la différence quadratique moyenne (*root-mean-square* en anglais : RMS) entre deux contours, aussi proposée par (Hermes, 1998). C'est

---

<sup>15</sup> Ici, il s'agit d'une mesure de convergence de deux contours intonatifs : plus deux contours sont similaires, plus la valeur de corrélation augmente. Nous utilisons par commodité le terme de « distance » pour indiquer la proximité entre deux courbes intonatives.



une mesure de la distance de registre de F0 entre deux courbes intonatives. La formule utilisée ici pour calculer la différence RMS est une adaptation de la formule donnée par d'Alessandro et al. (2011) :

$$R = \sqrt{\frac{\sum_i w(i)((f_1(i) - m_1) - (f_2(i) - m_2))^2}{\sum_i w(i)}}$$

La différence RMS mesure la différence moyenne en demi-tons entre deux contours intonatifs. Elle est toujours positive avec la valeur 0 pour des courbes identiques. Elle diffère de la corrélation qui mesure la similarité entre les formes de deux courbes intonatives : cette différence RMS est une mesure de dissimilarité qui représente l'aire de l'espace entre les deux courbes.

#### 2.4.5.2. Classement des affects sociaux

Par la suite, une analyse de classification hiérarchique a été effectuée respectivement sur les corrélations et les RMS entre les contours intonatifs des affects sociaux pour chaque longueur de l'énoncé.

- Stimuli d'une syllabe

Les 19 affects sociaux produits sur les stimuli monosyllabiques sont regroupés en fonction de la corrélation et de la distance du RMS de F0 entre deux contours intonatifs. L'énoncé utilisé est « shu4 », « arbre » en français. Les résultats sont représentés dans la figure 17:

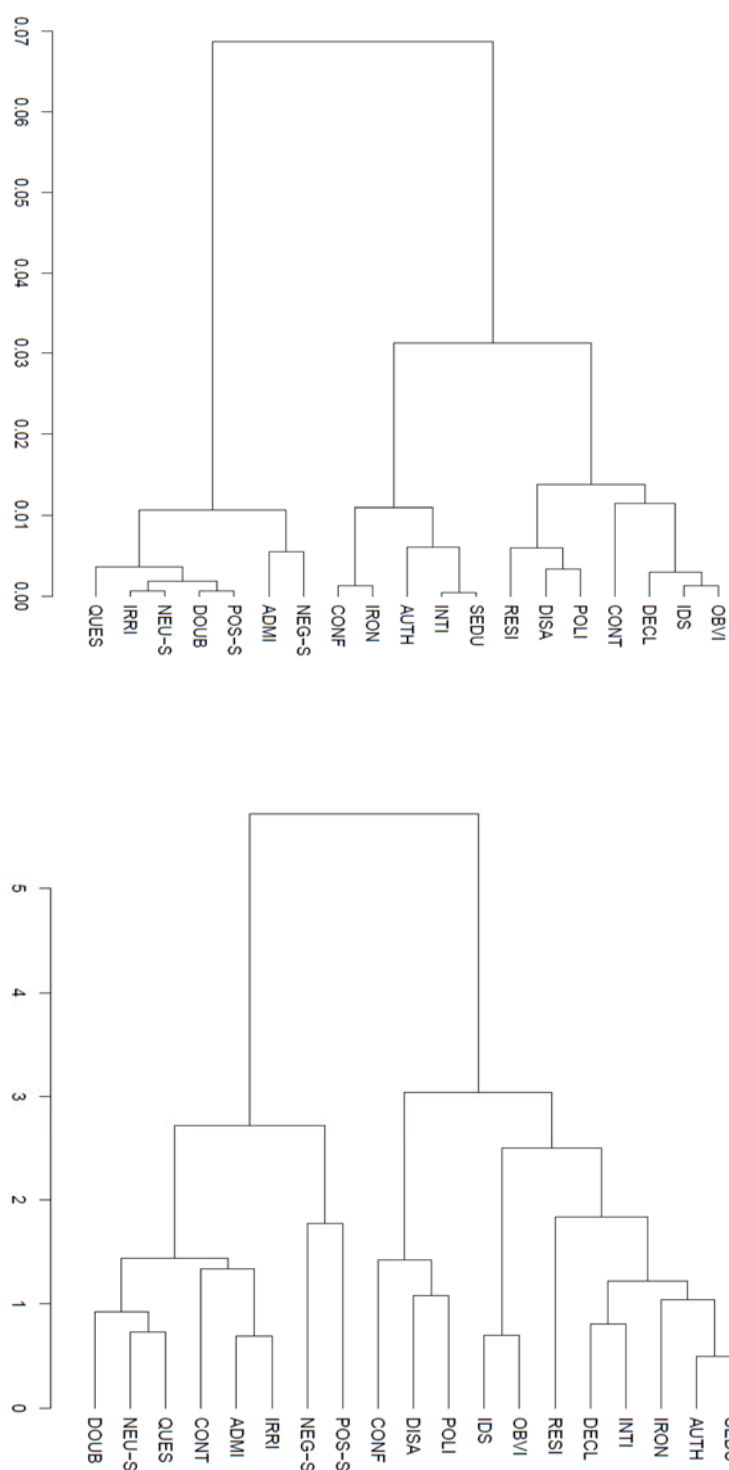


Figure 17 : Dendrogrammes montrant les distances entre les affects sociaux exprimés avec un mot monosyllabique basées sur les coefficients de corrélation pondérés (en haut) et sur la différence moyenne quadratique du registre de F0 (en bas) entre les contours intonatifs des affects sociaux.

La classification hiérarchique permet d'extraire, au plus haut niveau de différenciation, trois catégories pour les 19 affects sociaux sur le plan de la corrélation de la forme des

contours de F0. La première catégorie contient la *question* (QUES), l'*irritation* (IRRI), l'exclamation de *surprise neutre* (NEU-S), le *doute* (DOUB), l'exclamation de *surprise positive* (POS-S), l'*admiration* (ADMI) et l'exclamation de *surprise négative* (NEG-S). La deuxième catégorie est composée de la *confiance* (CONF), l'*ironie* (IRON), l'*autorité* (AUTH), l'*intimité* (INTI) et la *séduction* (SEDU). La troisième catégorie regroupe la *résignation* (RESI), la *déception* (DISA), la *politesse* (POLI), le *mépris* (CONT), la *déclaration* (DECL), le *maternel* (IDS) et l'*évidence* (OBVI). Cependant, les affects sociaux montrent de nombreuses similarités entre eux, en dépit des différences de catégories.

Sur la base de la différence des moindres carrés, les affects sociaux sont classés en deux catégories : l'une contient la *confiance* (CONF), l'*ironie* (IRON), l'*autorité* (AUTH), l'*intimité* (INTI), la *séduction* (SEDU), la *résignation* (RESI), la *déception* (DISA), la *politesse* (POLI), la *déclaration* (DECL), le *maternel* (IDS) et l'*évidence* (OBVI) ; l'autre regroupe les trois expressions d'exclamation de *surprise*, le *doute*, la *question*, l'*admiration*, le *mépris* et l'*irritation*.

L'expression du *mépris* (CONT) est regroupée différemment selon la corrélation et le RMS de F0 de contour intonatif : elle est plus proche de l'*admiration*, de l'*irritation* et des attitudes interrogatives, inattendues et incertaines (la *question*, les trois expressions de *surprise* et le *doute*) en terme de hauteur de F0, mais des attitudes assertives (par exemple la *déclaration* et l'*évidence*) et des affects montrant la relation sociale interpersonnelle et le contexte communicatif (par exemple le *maternel* et la *politesse*) en terme de forme de contour. La figure 18 présente la comparaison de la courbe intonative du *mépris* avec celle de l'exclamation de *surprise neutre*, de la *question*, de la *déclaration* et de la *politesse*.

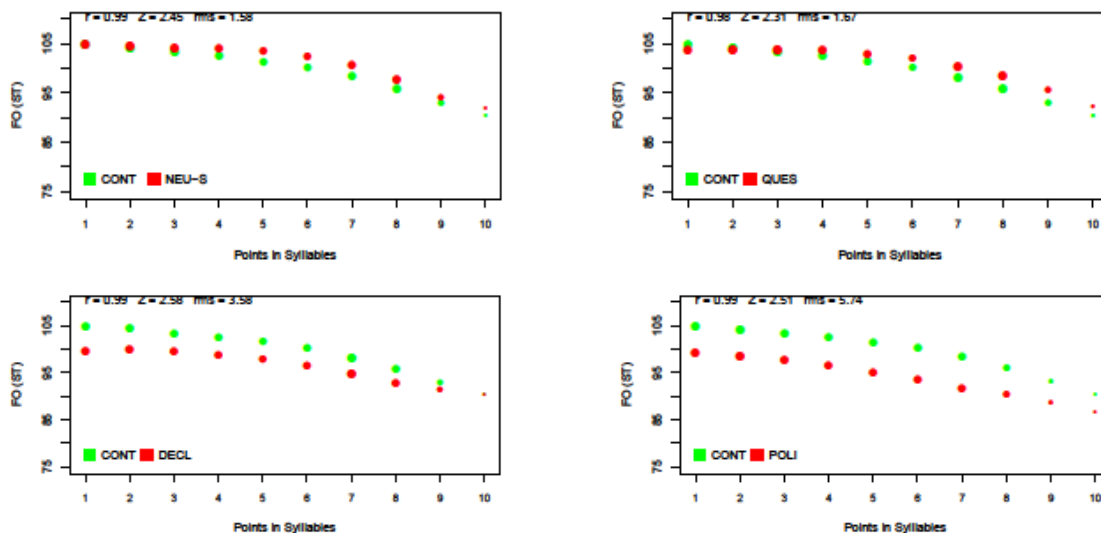


Figure 18 : Comparaison de la courbe intonative du *mépris* avec celle de l'exclamation de *surprise neutre*, de la *question*, de la *déclaration* et de la *politesse*. Les valeurs de F0 des contours sont représentées en demi-ton et normalisés à 10 points par voyelle.  $r$  ( $-1 < r < 1$ ): coefficient de corrélation pondéré entre deux courbes intonatives;  $Z$ : score de la transformation de Fisher;  $rms$  ( $rms > 0$ ): différence moyenne quadratique entre deux courbes intonatives.

- Stimuli de deux syllabes

Le résultat de la classification hiérarchique pour les énoncés de deux syllabes est représenté par la figure 19. L'énoncé utilisé est « fang4 xue2 », « l'école est finie » en français.

En matière de corrélation de la forme de contour de F0, les affects sociaux sont classés, au plus haut niveau, en trois catégories. La première catégorie regroupe les trois expressions de l'exclamation de *surprise*, la *question* (QUES), l'*admiration* (ADMI), la *séduction* (SEDU), l'*intimité* (INTI) et l'*ironie* (IRON); la deuxième catégorie contient la *déclaration* (DECL), la *confiance* (CONF), la *politesse* (POLI), l'*évidence* (OBVI), l'*autorité* (AUTH), l'*irritation* (IRRI), la *déception* (DISA) et la *résignation* (RESI); la troisième catégorie est composée du *maternel* (IDS), du *doute* (DOUB) et du *mépris* (CONT).

En ce qui concerne la hauteur de F0 de l'énoncé, les affects sociaux sont regroupés en deux catégories. La première catégorie rassemble l'exclamation de *surprise négative* (NEG-S) et *surprise neutre* (NEU-S), la *question* (QUES), l'*admiration* (ADMI), l'*irritation* (IRRI), l'*intimité* (INTI) et l'*ironie* (IRON), ainsi que l'exclamation de *surprise positive* (POS-S) qui est isolée des autres affects sociaux du groupe à cause de son registre de F0 plus haut. La seconde catégorie contient le reste des affects sociaux.

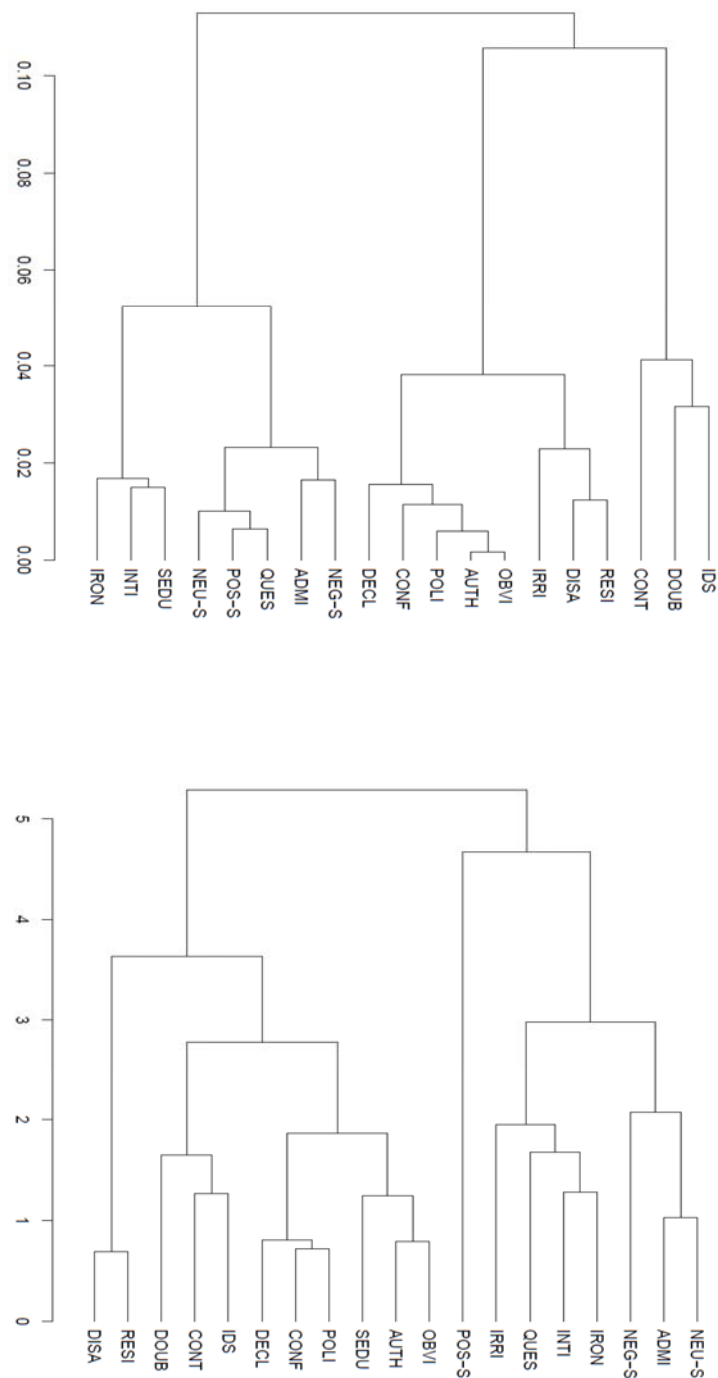


Figure 19 : Dendrogrammes montrant les distances entre les affects sociaux exprimés avec un mot dissyllabique basées sur les coefficients de corrélation pondérés (en haut) et sur la différence moyenne quadratique du registre de F0 (en bas) entre les contours intonatifs des affects sociaux.

La *séduction*, de même que l'*irritation*, sont classées de manière différente selon la forme de son contour intonatif et la hauteur de F0. La *séduction* se montre plus proche des attitudes d'assertion (par exemple la *déclaration* et l'*évidence*) et des affects sociaux

indiquant la relation sociale interpersonnelle (l'*autorité* et la *politesse*) au niveau de la hauteur de F0, mais plus proche de l'*intimité*, de l'*ironie*, ainsi que des attitudes interrogatives et inattendues (la *question* et les trois expressions de *surprise*) au niveau de la forme. L'*irritation* s'est située avec les attitudes interrogatives et inattendues pour une similarité en registre, mais avec les attitudes assertives à cause de la forme de contour intonatif similaire. La figure 20 présente la comparaison de la courbe intonative de la *séduction* et l'*irritation* avec celles de la *déclaration*, de la *politesse*, de l'exclamation de *surprise neutre* et de la *question*.

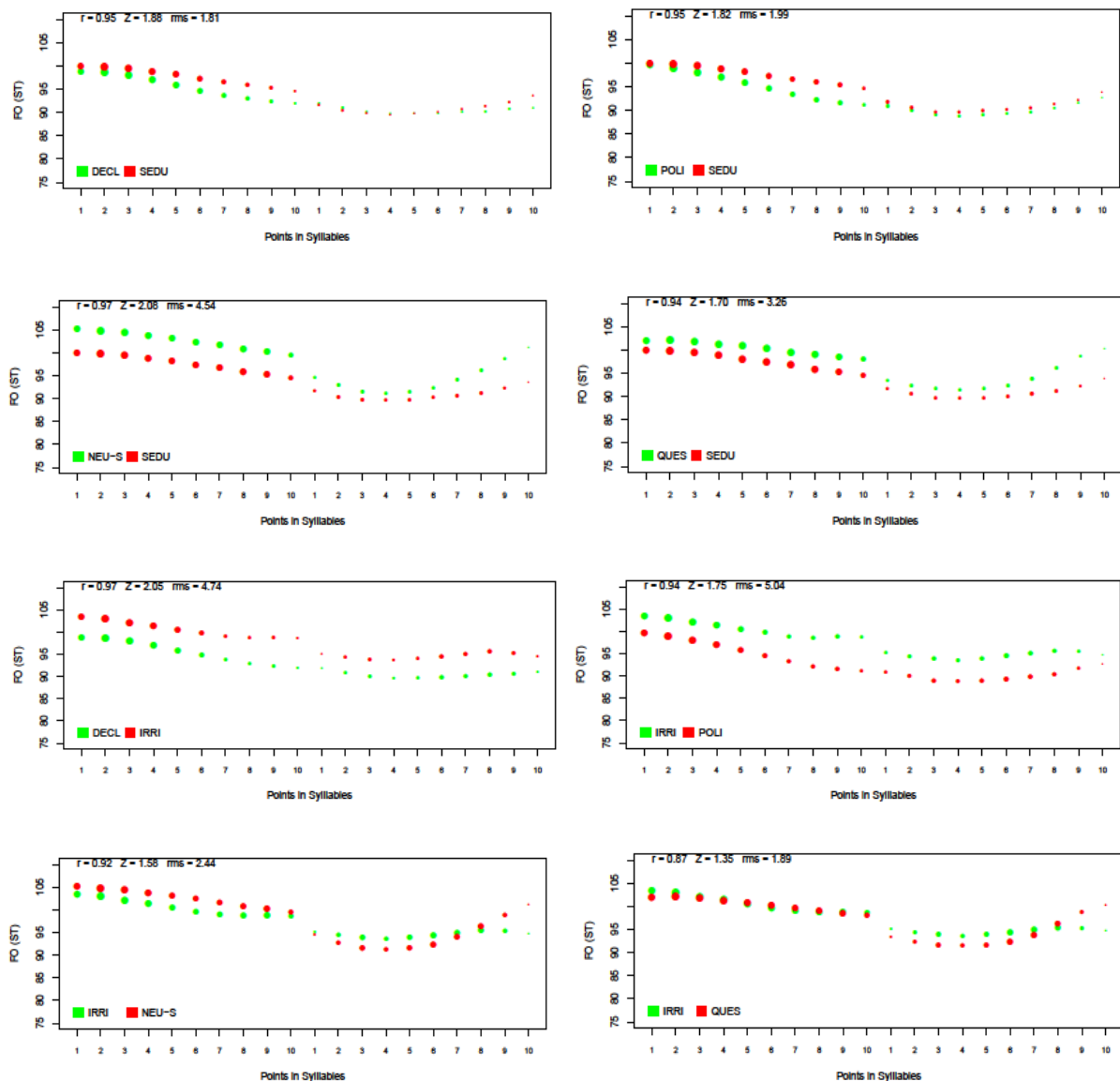


Figure 20 : Comparaison de la courbe intonative de la séduction et de l'irritation avec celle de l'exclamation de surprise neutre, de la question, de la déclaration et de la politesse. Les valeurs de F0 des contours sont représentées en demi-ton et normalisés à 10 points par voyelle.  $r$  ( $-1 < r < 1$ ): coefficient de corrélation pondéré entre deux courbes intonatives;  $Z$ : score de la transformation de Fisher;  $rms$  ( $rms > 0$ ): différence moyenne quadratique entre deux courbes intonatives.

Par ailleurs, le *maternel*, le *doute* et le *mépris* se montrent similaires en RMS aussi bien qu'à la forme. Cependant, bien qu'ils soient proches des attitudes assertives et des affects sociaux comme la *séduction* et l'*autorité* au niveau du registre, ils se différencient d'eux par leur forme de contour intonatif : la première syllabe de l'énoncé de ces trois affects sociaux commence avec un registre plus haut et forme une pente descendante plus importante ; en même temps, ils montrent une montée plus importante à la deuxième moitié de la deuxième syllabe. La figure 21 présente quelques exemples de la comparaison de la courbe prosodique du *maternel*, de *doute* et du *mépris* avec celles de la *déclaration* et de la *politesse*.

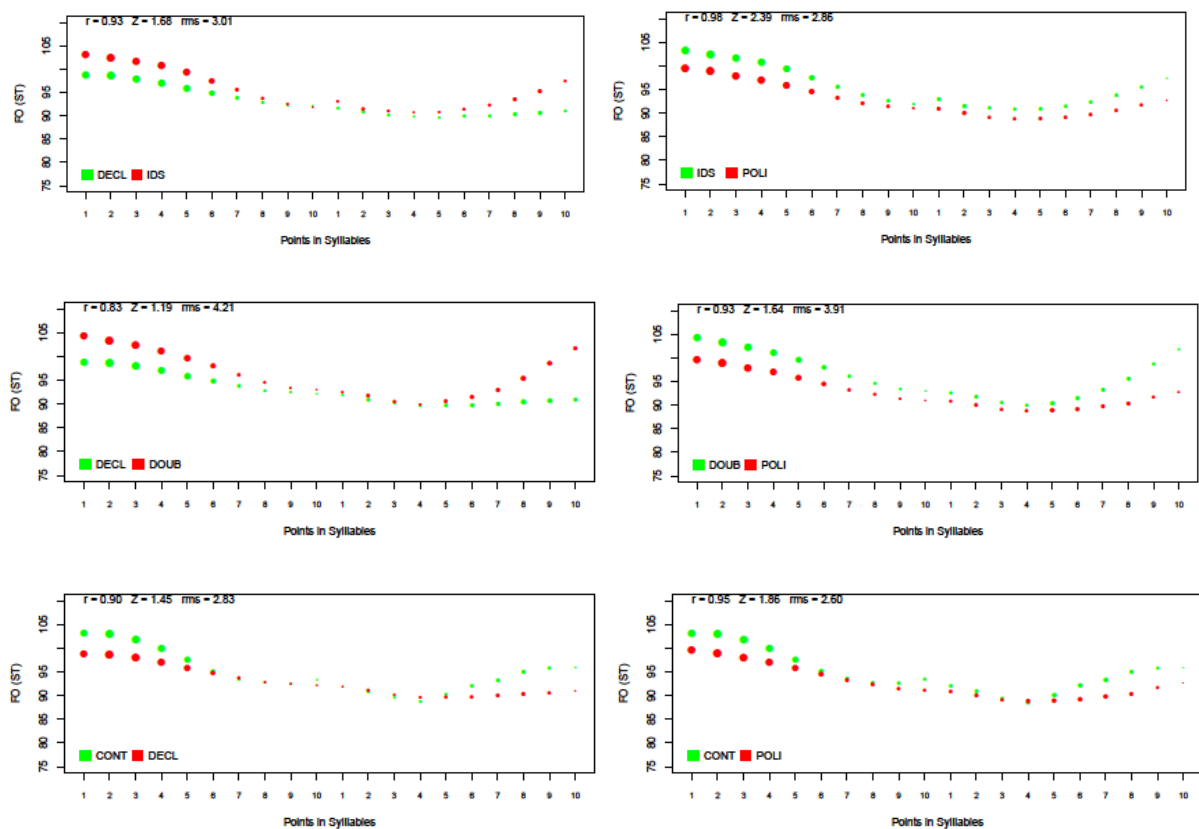


Figure 21 : Comparaison de la courbe intonative du *maternel*, du *doute* et du *mépris* avec celle de la *déclaration* de la *politesse*. Les valeurs de F0 des contours sont représentées en demi-ton et normalisés à 10 points par voyelle.  $r$  ( $-1 < r < 1$ ): coefficient de corrélation pondéré entre deux courbes intonatives;  $Z$  : score de la transformation de Fisher ;  $rms$  ( $rms > 0$ ): différence moyenne quadratique entre deux courbes intonatives.

- Stimuli de quatre syllabes

Le résultat de l'analyse de classification hiérarchique pour les énoncés de quatre syllabes est illustré par les dendrogrammes de la figure 22. La phrase utilisée est « zhang1 yi1 sheng1 lai2 » dont la traduction française est « Docteur Zhang viendra ».

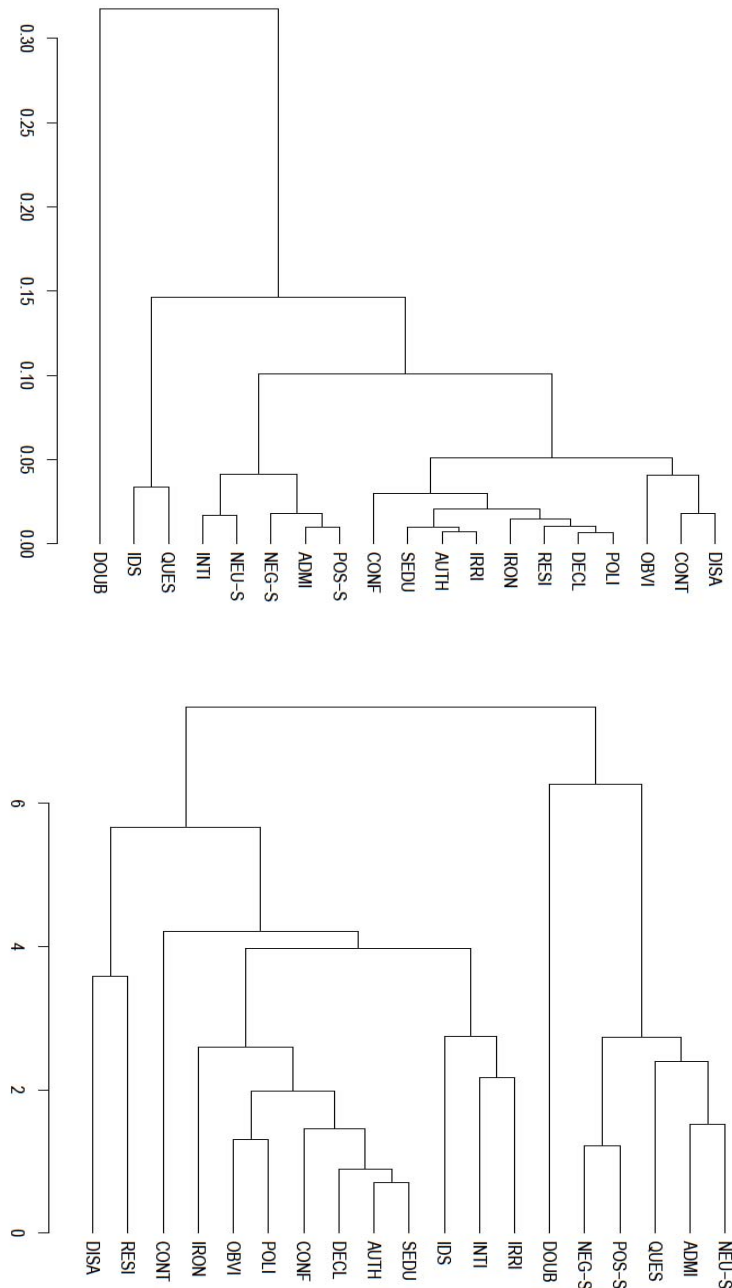


Figure 22 : Dendrogrammes montrant les distances entre les affects sociaux exprimés avec une phrase de quatre syllabes basées sur les coefficients de corrélation pondérés (en haut) et sur la différence moyenne quadratique du registre de F0 (en bas) entre les contours intonatifs des affects sociaux.



En ce qui concerne la corrélation de la forme de contour intonatif, le *doute* (DOUB) est l'attitude la plus distincte des autres, puis suivent le *maternel* (IDS) et la *question* (QUES). Les autres affects sociaux se répartissent en deux catégories : les trois expressions de l'exclamation de *surprise*, l'*intimité* (INTI) et l'*admiration* (ADMI) dans la première catégorie et le reste des affects sociaux dans la seconde.

Le *doute* se montre différent par rapport aux autres affects sociaux au niveau de la forme de son contour intonatif, du fait que les valeurs de F0 de la 4<sup>e</sup> syllabe de l'énoncé sont marquées par une qualité de voix craquée (*creaky voice* en anglais) et montrent donc des valeurs beaucoup plus basses que celles des autres affects sociaux. La forme du contour intonatif de l'expression de *doute* ressemble à celle des attitudes interrogatives et inattendues, notamment à celle de la *question* si l'on fait abstraction du contour local de la dernière syllabe. Le contour intonatif de l'expression *maternelle* montre une déformation au niveau de la 4<sup>e</sup> syllabe, mais pour le reste, la forme de sa courbe intonative est très similaire avec celle de la *question*.

En regardant les contours intonatifs de ces trois affects sociaux, nous remarquons que sur leur dernière syllabe, la première moitié suit une pente descendante puis remonte, cela forme un contour descendant-remontant, au lieu d'un contour montant correspondant au ton lexical de la 4<sup>e</sup> syllabe. Ce changement est dû à la coarticulation avec ton précédent qui est ici un ton plat et haut. D'après Lin & Yan (1992), l'effet coarticulatoire est généralement unidirectionnel, et c'est toujours le ton précédent qui influence le ton suivant. D'après les auteurs, la coarticulation tonale en chinois mandarin est souvent à l'origine de la perturbation de F0 dans l'énoncé. Le même phénomène a été également observé pour les contours intonatifs des trois expressions d'exclamation de *surprise*, d'*intimité* (INTI) et d'*admiration* (ADMI).

Inversement, les attitudes assertives subissent en général une déclinaison globale sur l'ensemble de l'énoncé, même si le ton terminal est un ton montant. Leur mouvement prosodique montre peu de changement même en fin d'énoncé. La figure 23 présente les contours intonatifs de ces affects sociaux, ainsi que ceux de l'exclamation de *surprise neutre* et de *surprise positive*, de la *déclaration* et de la *politesse*.

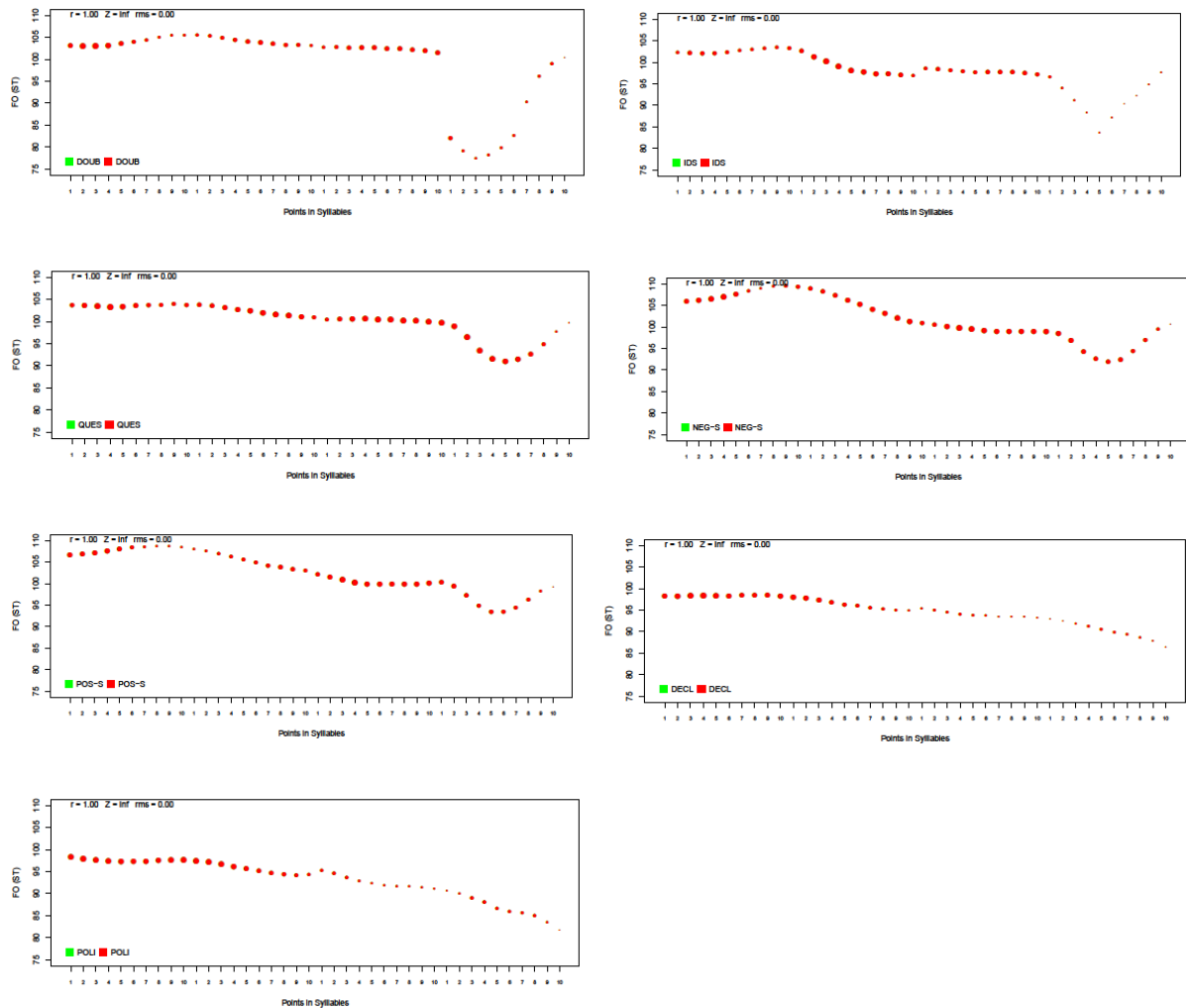


Figure 23 : Les courbes intonatives du *doute*, du *maternel*, de la *question*, de l'exclamation de *surprise positive* et de *surprise négative*, de la *déclaration* et de la *politesse*. Les valeurs de F0 des contours sont représentées en demi-ton et normalisés à 10 points par voyelle.  $r$  ( $-1 < r < 1$ ): coefficient de corrélation pondéré entre deux courbes intonatives;  $Z$ : score de la transformation de Fisher;  $rms$  ( $rms > 0$ ): différence moyenne quadratique entre deux courbes intonatives.

En matière de RMS de F0, les 19 affects sociaux sont regroupés en deux catégories principales : une catégorie est composée de l'*admiration* et des attitudes interrogatives, inattendues et incertaines ; l'autre catégorie contient le reste des affects sociaux. A l'intérieur de la première catégorie, le *doute* est isolé des autres affects sociaux. Cela pourrait être dû au fait que les valeurs de F0 de la dernière syllabe de l'énoncé sont beaucoup plus petites que celles des autres affects sociaux du groupe à cause de la présence de la qualité de voix craquée. Au sein de la deuxième catégorie, la *déception* et la *résignation* se distinguent des autres affects sociaux du groupe, à cause d'un registre de F0 beaucoup plus bas.

L'*intimité* est classée différemment selon des critères de forme de contour intonatif ou de hauteur de F0. Elle se montre proche de l'*irritation*, du *maternel* et des attitudes d'assertion au niveau du registre, mais proche des expressions d'exclamation de *surprise* et de l'*admiration* au niveau de la forme. La figure 24 présente une comparaison de la courbe intonative de l'*intimité* avec celle de la *déclaration*, de la *politesse*, de l'exclamation de *surprise neutre* et de *surprise négative*.

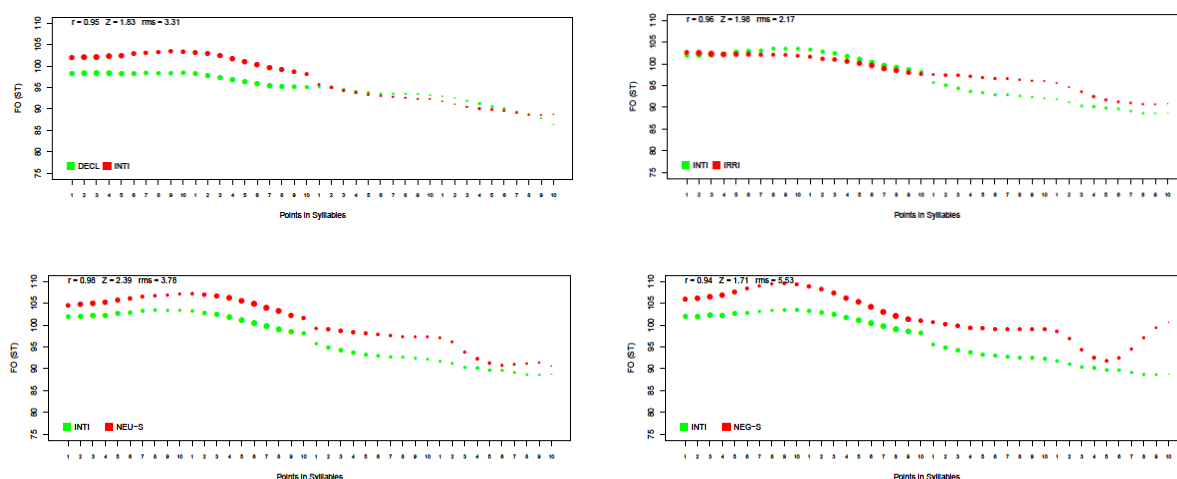


Figure 24 : Comparaison de la courbe intonative de l'*intimité* avec celle de la *déclaration*, de l'*irritation*, de l'exclamation de *surprise neutre* et de *surprise négative*. Les valeurs de F0 des contours sont représentées en demi-ton et normalisés à 10 points par voyelle.  $r$  ( $-1 < r < 1$ ): coefficient de corrélation pondéré entre deux courbes intonatives;  $Z$ : score de la transformation de Fisher;  $rms$  ( $rms > 0$ ): différence moyenne quadratique entre deux courbes intonatives.

Par ailleurs, comme au niveau de la forme de contour intonatif, l'expression de *mépris* est aussi regroupée avec les attitudes assertives et les affects sociaux indiquant la relation sociale interpersonnelle au niveau de la hauteur. Cependant, elle se différencie des autres affects sociaux du groupe au niveau du registre. En regardant son contour de F0 et en écoutant le stimulus, nous remarquons une voix grave et craquée sur la dernière syllabe de l'énoncé. Les valeurs de F0 de la dernière syllabe sont très basses par rapport à celles des autres affects sociaux du groupe. Quelques exemples de comparaison entre la forme de contour intonatif du *mépris* et celles de la *déclaration* et de la *politesse* sont illustrés dans la figure 25.

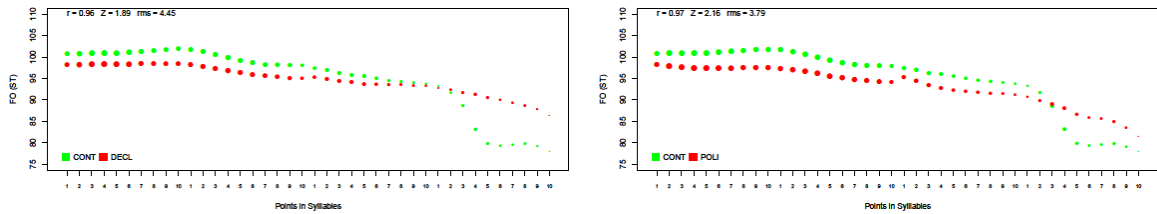
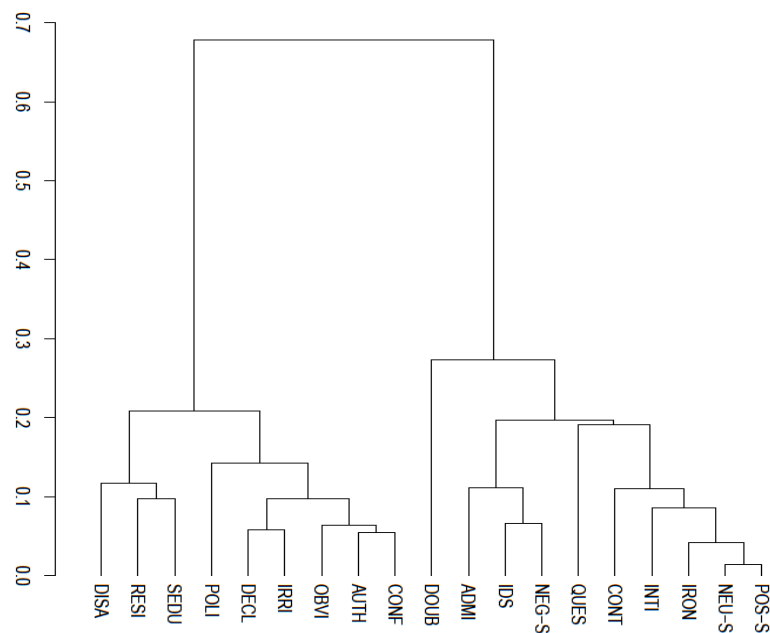


Figure 25 : Comparaison de la courbe intonative du *mépris* avec celle de la *déclaration* et de la *politesse*. Les valeurs de F0 des contours sont représentées en demi-ton et normalisés à 10 points par voyelle.  $r$  ( $-1 < r < 1$ ): coefficient de corrélation pondéré entre deux courbes intonatives;  $Z$ : score de la transformation de Fisher ;  $rms$  ( $rms > 0$ ): différence moyenne quadratique entre deux courbes intonatives.

- Stimuli de neuf syllabes

Le résultat de la classification hiérarchique pour les énoncés de neuf syllabes est illustré par les dendrogrammes dans la figure 26. La phrase utilisée est « wang2 yi1 sheng1 ta1 san1 gu1 ma1 xiu1 jia4 » dont la traduction française est « La troisième tante paternelle de Docteur Wang prend un congé ».



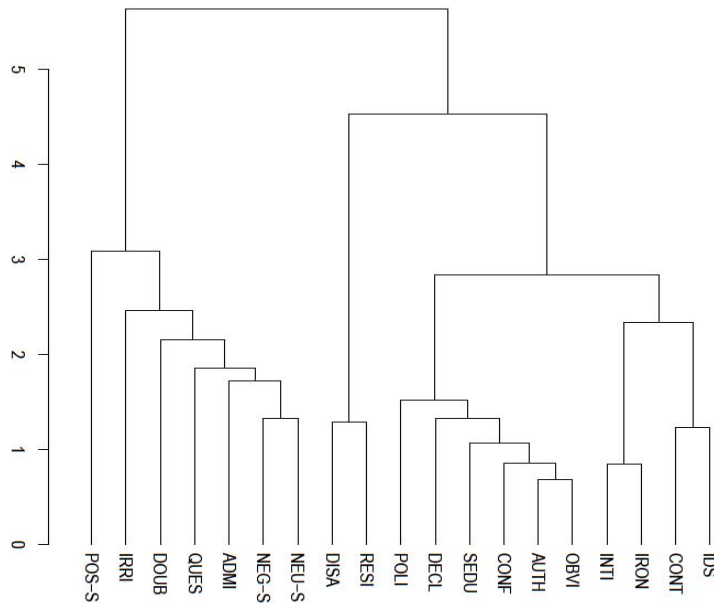


Figure 26 : Dendrogrammes montrant les distances entre les affects sociaux exprimés avec une phrase de neuf syllabes basées sur les coefficients de corrélation pondérés (en haut) et sur la différence moyenne quadratique du registre de F0 (en bas) entre les contours intonatifs des affects sociaux.

Concernant la forme du contour intonatif, les affects sociaux sont divisés en deux groupes principaux. Un des groupes rassemble la *déception* (DISA), la *résignation* (RESI), la *séduction* (SEDU), la *politesse* (POLI), la *déclaration* (DECL), l'*irritation* (IRRI), l'*évidence* (OBVI), l'*autorité* (AUTH) et la *confiance* (CONF) ; alors que l'autre regroupe la *question* (QUES), le *doute* (DOUB) et les trois expressions de l'exclamation de *surprise*, avec l'*ironie* (IRON), l'*intimité* (INTI), le *mépris* (CONT) et le *maternel* (IDS).

En matière de RMS de F0 de l'énoncé, trois principales catégories d'affects sociaux sont observées. La *déception* et la *résignation* sont isolées des autres affects sociaux pour leur registre de F0 plus bas (voir figure 29); le reste des affects sociaux est réparti en deux catégories : l'une est composée des attitudes assertives et des affects sociaux montrant la relation sociale interpersonnelle et le contexte social de l'interaction, l'autre regroupe les affects sociaux interrogatifs, inattendus et incertains avec l'*irritation* et l'*admiration*.

L'*ironie*, l'*intimité*, le *mépris* et le *maternel* sont regroupés avec les attitudes assertives au niveau du registre de F0 de l'énoncé, mais avec les attitudes interrogatives, inattendues et incertaines au niveau de la forme de courbe prosodique. En confrontant leurs contours intonatifs avec ceux des attitudes assertives, nous observons que généralement les contours intonatifs de ces quatre affects sociaux commencent dans un registre de F0

plus bas, puis montent progressivement jusqu'à la première moitié de la dernière syllabe, et se terminent avec une descente assez importante sur la dernière syllabe de l'énoncé. Ce mouvement intonatif est aussi observé sur les contours de F0 des attitudes interrogatives, inattendues et incertaines. Au contraire, les contours intonatifs des autres attitudes assertives restent toujours très plats et descendent en fin d'énoncé.

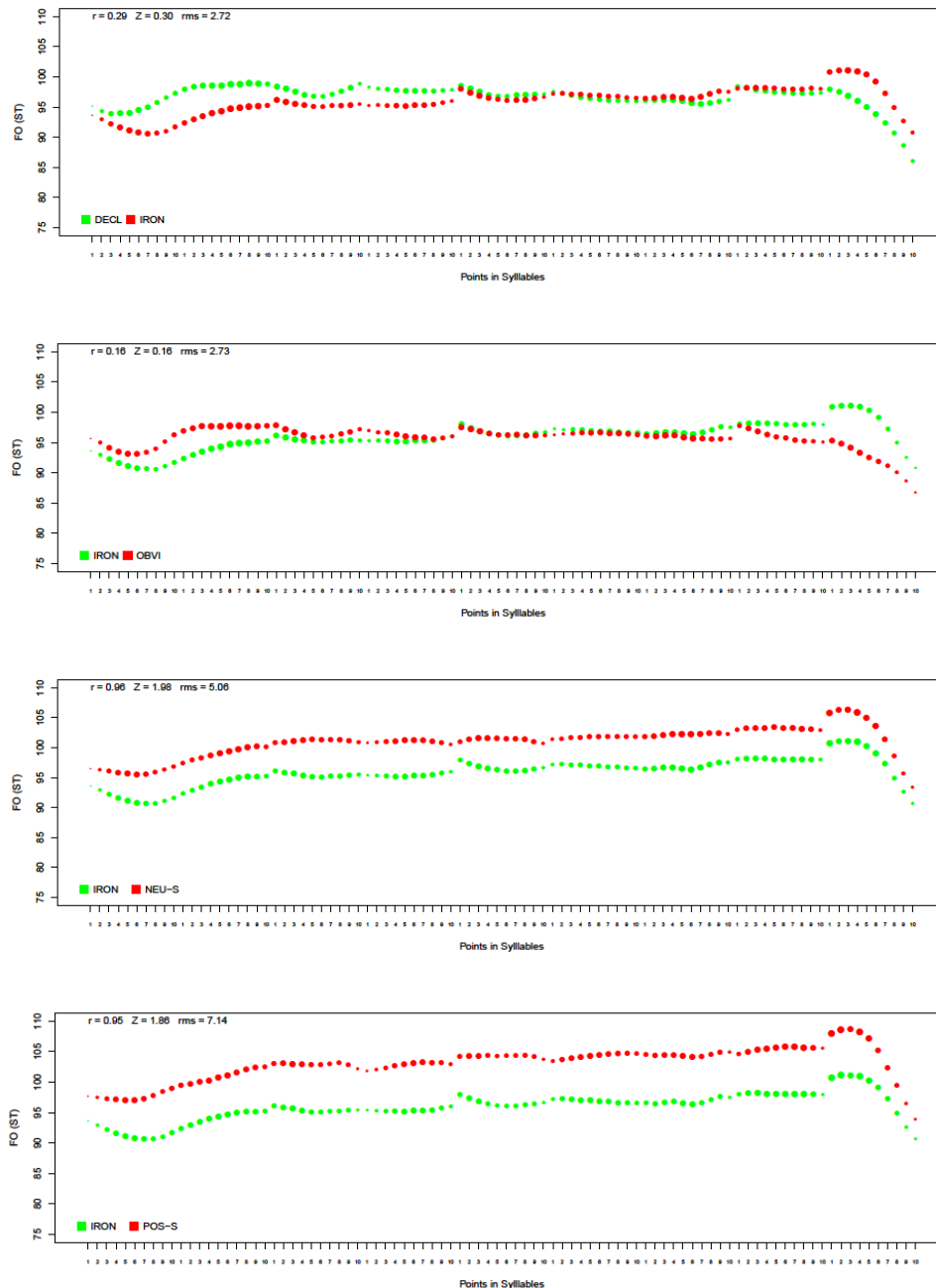
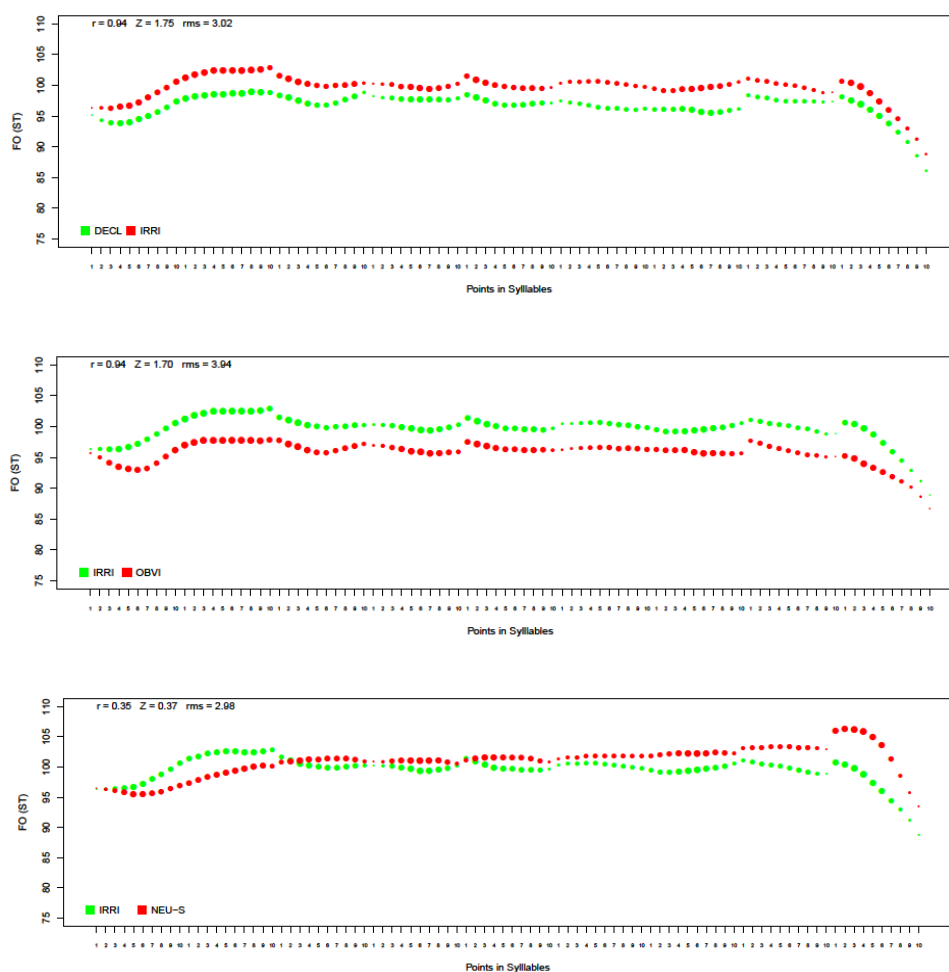


Figure 27 : Comparaison de la courbe intonative l'ironie avec celle de la *déclaration*, de l'*évidence*, de l'exclamation de *surprise neutre* et de *surprise positive*. Les valeurs de F0 des contours sont représentées en demi-ton et normalisés à 10 points par voyelle.  $r$  ( $-1 < r < 1$ ): coefficient de corrélation pondéré entre deux courbes intonatives;  $Z$ : score de la transformation de Fisher;  $rms$  ( $rms > 0$ ): différence moyenne quadratique entre deux courbes intonatives.

La figure 27 donne des exemples de comparaison entre la forme du contour intonatif de l'*ironie* et celles de la *déclaration*, de l'*évidence*, de l'exclamation de *surprise neutre* et de *surprise positive*.

L'*irritation* est classée avec les attitudes interrogatives, inattendues et incertaines au plan du registre, mais avec les attitudes assertives au plan de la forme de contour intonatif. A l'opposé de la *question*, du *doute* et des expressions de *surprise*, dont le contour intonatif monte progressivement sur l'ensemble de l'énoncé, le contour de l'*irritation* montre une déclinaison générale à l'énoncé entier. De l'autre côté, en comparaison des attitudes assertives, l'*irritation* a un registre plus élevé. Quelques exemples de comparaison entre la courbe intonative de l'*irritation* et celles de la *déclaration*, de l'*évidence*, de l'exclamation de *surprise neutre* et de la *question* sont illustrés par la figure 28.



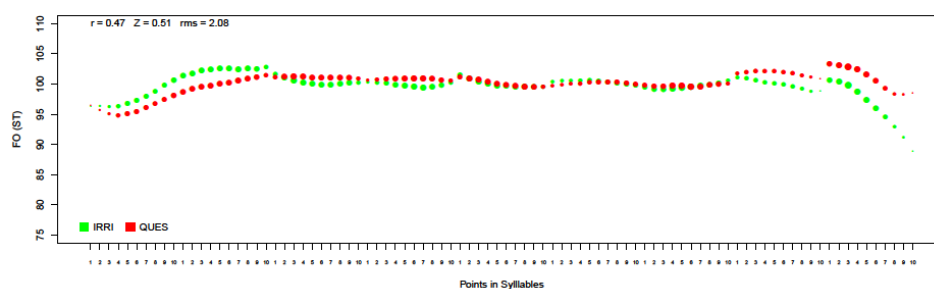
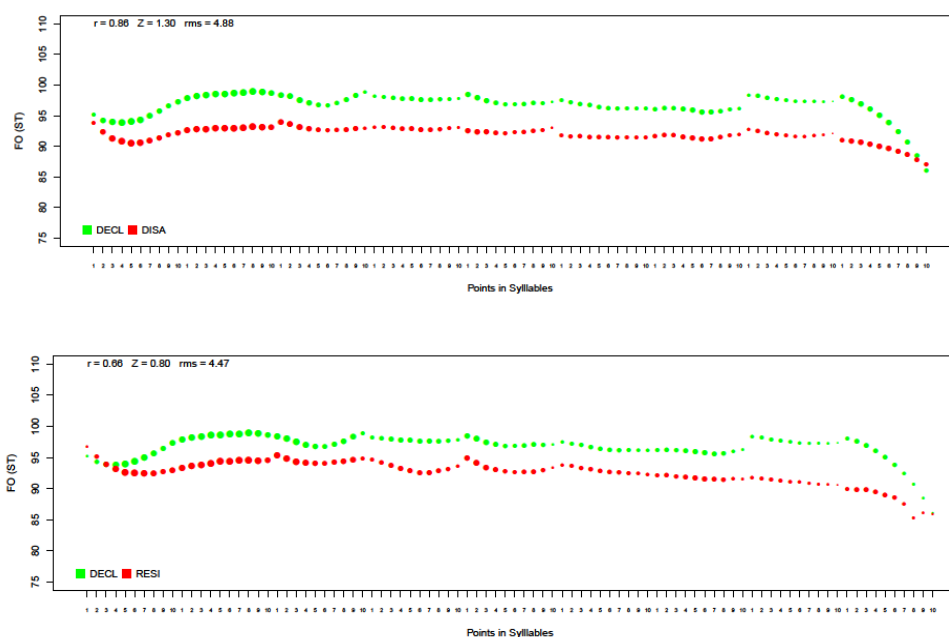


Figure 28 : Comparaison de la courbe intonative de l'*irritation* avec celle de la *déclaration*, de l'*évidence*, de l'exclamation de *surprise neutre* et de la *question*. Les valeurs de F0 des contours sont représentées en demi-ton et normalisés à 10 points par voyelle.  $r$  ( $-1 < r < 1$ ): coefficient de corrélation pondéré entre deux courbes intonatives;  $Z$ : score de la transformation de Fisher;  $rms$  ( $rms > 0$ ): différence moyenne quadratique entre deux courbes intonatives.

Enfin, la *résignation* et la *déception* sont proches des attitudes assertives et des affects sociaux montrant la relation sociale interpersonnelle et le contexte social de la communication au niveau du contour intonatif, mais leur registre est beaucoup plus bas que celui de ces affects sociaux. L'expression de *doute* est regroupée avec les attitudes interrogatives et inattendues dans les deux domaines, mais elle se montre un peu différente en terme de contour de F0: son contour intonatif est caractérisé par deux pics de F0 respectivement au début (la 2<sup>e</sup> syllabe) et à la fin (la dernière syllabe) de l'énoncé; or, le contour intonatif des attitudes interrogatives ne montre qu'un seul pic sur la dernière syllabe. Ces observations sont illustrées par la figure 29.





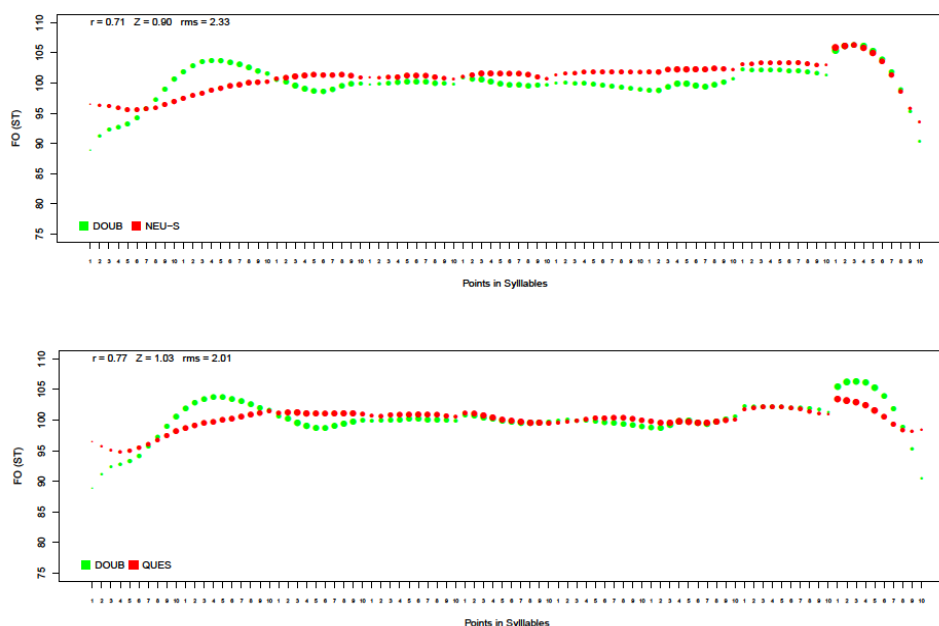


Figure 29 : Contours intonatifs de la *déception* et de la *résignation* comparés avec celui de la *déclaration*, ainsi que le contour intonatif du *doute* comparé avec celui de l'exclamation de *surprise neutre* et de la *question*. Les valeurs de F0 des contours sont représentées en demi-ton et normalisés à 10 points par voyelle.  $r$  ( $-1 < r < 1$ ): coefficient de corrélation pondéré entre deux courbes intonatives;  $Z$ : score de la transformation de Fisher ;  $rms$  ( $rms > 0$ ): différence moyenne quadratique entre deux courbes intonatives.

#### 2.4.5.3. Confusion perçue des affects sociaux

Ayant examiné les similarités et les dissemblances des contours intonatifs des 19 affects sociaux, nous poursuivons à comparer la similarité des affects sociaux au niveau du mouvement intonatif avec leur similarité au niveau de la perception auditive. L'objectif de cette analyse est de donner des explications possibles aux confusions perceptives entre les affects sociaux obtenues dans le test perceptif (voir partie 2.3.).

La table 12 présente les matrices de confusion des 19 affects sociaux chinois pour les quatre énoncés de 1, 2, 4 et 9 syllabes.

En observant les confusions entre les affects sociaux, dans un premier temps, nous remarquons une attractivité importante de la *déclaration* (DECL) sur les autres affects sociaux quand l'énoncé est court (une et deux syllabes). Il se peut que ces énoncés courts ne permettent pas à la prosodie seule la transmission de suffisamment d'informations affectives, les auditeurs distinguent donc mal ces affects sociaux. Mais cette hypothèse reste à vérifier avec les autres énoncés du corpus utilisé dans l'expérience perceptive.

1syllabe	ADMI	AUTH	CONF	DECL	DISA	DOUB	IRRI	IRON	INTI	CONT	IDS	OBVI	POLI	QUES	RESI	SEDU	NEG S	NEU S	POS S
ADMI	13,79	0	10,34	20,69	0	0	3,45	0	3,45	6,9	3,45	24,14	0	3,45	0	0	0	10,34	0
AUTH	0	16,67	6,67	26,67	3,33	0	23,33	0	3,33	3,33	0	16,67	0	0	0	0	0	0	0
CONF	0	0	6,67	46,67	6,67	0	0	0	3,33	0	3,33	10	13,33	0	10	0	0	0	0
DECL	0	0	10	53,33	0	0	0	0	3,33	0	0	26,67	6,67	0	0	0	0	0	0
DISA	0	6,67	0	33,33	16,67	0	0	0	3,33	0	0	10	3,33	0	23,33	0	3,33	0	0
DOUB	10	0	3,33	6,67	3,33	13,33	0	3,33	23,33	3,33	0	10	3,33	10	3,33	0	3,33	3,33	0
IRRI	0	6,67	10	0	0	0	60	0	6,67	6,67	0	10	0	0	0	0	0	0	0
IRON	0	0	3,33	16,67	6,67	0	3,33	3,33	6,67	46,67	0	0	3,33	0	10	0	0	0	0
INTI	0	0	13,33	26,67	0	0	0	6,67	16,67	10	3,33	13,33	6,67	0	3,33	0	0	0	0
CONT	3,33	3,33	6,67	30	0	10	3,33	0	6,67	0	0	33,33	0	0	0	0	3,33	0	0
IDS	0	0	6,67	20	0	0	0	0	6,67	0	46,67	6,67	6,67	0	3,33	3,33	0	0	0
OBVI	0	6,67	13,33	33,33	3,33	0	6,67	3,33	0	0	0	26,67	3,33	0	3,33	0	0	0	0
POLI	3,33	0	6,67	73,33	0	0	0	0	0	0	10	0	6,67	0	0	0	0	0	0
QUES	3,33	0	3,33	26,67	0	3,33	0	6,67	0	3,33	16,67	6,67	13,33	0	0	0	0	6,67	10
RESI	0	0	6,67	36,67	16,67	0	3,33	0	6,67	6,67	0	3,33	0	0	20	0	0	0	0
SEDU	0	0	6,67	43,33	3,33	0	0	0	3,33	0	10	0	13,33	0	13,33	6,67	0	0	0
NEG-S	0	6,67	6,67	0	0	6,67	16,67	3,33	0	0	0	20	0	0	0	3,33	3,33	20	13,33
NEU-S	10	3,33	20	3,33	3,33	3,33	0	0	3,33	0	0	0	0	0	3,33	0	0	16,67	3,33
POS-S	13,33	6,67	0	3,33	3,33	3,33	10	0	23,33	0	6,67	0	3,33	0	0	0	0	3,33	23,33

2syllables	ADMI	AUTH	CONF	DECL	DISA	DOUB	IRRI	IRON	INTI	CONT	IDS	OBVI	POLI	QUES	RESI	SEDU	NEG S	NEU S	POS S
ADMI	23,33	0	6,67	13,33	3,33	6,67	0	0	10	0	6,67	0	13,33	3,33	0	0	0	10	3,33
AUTH	0	50	6,67	26,67	3,33	0	10	0	0	0	0	3,33	0	0	0	0	0	0	0
CONF	0	13,33	0	70	6,67	0	0	0	0	0	6,67	3,33	0	0	0	0	0	0	0
DECL	0	10	3,33	63,33	3,33	0	0	0	0	0	0	10	6,67	0	3,33	0	0	0	0
DISA	0	0	0	16,67	70	0	0	0	0	0	3,33	0	0	0	10	0	0	0	0
DOUB	0	6,67	0	6,67	0	40	0	6,67	3,33	6,67	0	0	0	13,33	0	0	0	10	6,67
IRRI	0	26,67	13,33	13,33	0	0	26,67	0	6,67	0	0	10	0	0	3,33	0	0	0	0
IRON	0	10	0	23,33	3,33	6,67	0	10	0	26,67	0	3,33	3,33	3,33	10	0	0	0	0
INTI	10	10	3,33	13,33	3,33	10	0	3,33	10	0	3,33	3,33	3,33	16,67	0	0	0	6,67	3,33
CONT	0	3,33	0	13,33	0	6,67	0	30	3,33	26,67	0	3,33	6,67	0	3,33	0	0	3,33	0
IDS	0	0	0	10	0	0	0	0	3,33	0	66,67	0	6,67	0	0	13,33	0	0	0
OBVI	0	23,33	3,33	46,67	0	0	10	0	0	0	0	16,67	0	0	0	0	0	0	0
POLI	6,67	0	3,33	46,67	0	0	0	0	10	0	13,33	0	13,33	0	6,67	0	0	0	0
QUES	3,33	3,33	3,33	26,67	0	10	0	0	0	0	0	6,67	0	33,33	0	0	0	13,33	0
RESI	0	6,67	0	33,33	30	0	0	0	0	0	0	3,33	3,33	0	23,33	0	0	0	0
SEDU	0	0	0	10	0	0	0	0	13,33	0	40	0	20	0	0	16,67	0	0	0
NEG-S	3,33	3,33	0	3,33	0	33,33	6,67	13,33	0	6,67	0	0	0	0	0	0	20	10	0
NEU-S	3,33	3,33	6,67	0	0	26,67	0	10	0	3,33	0	13,33	0	16,67	0	0	0	10	6,67
POS-S	13,33	0	0	0	0	16,67	0	0	3,33	0	3,33	0	0	6,67	0	3,33	0	3,33	50

4syllables	ADMI	AUTH	CONF	DECL	DISA	DOUB	IRRI	IRON	INTI	CONT	IDS	OBVI	POLI	QUES	RESI	SEDU	NEG S	NEU S	POS S
ADMI	0	0	0	0	0	6,67	0	0	0	3,33	0	0	0	0	0	0	16,67	36,67	36,67
AUTH	0	53,33	0	10	3,33	0	13,33	3,33	0	3,33	0	10	0	3,33	0	0	0	0	0
CONF	3,33	6,67	6,67	63,33	3,33	0	0	0	0	0	0	6,67	10	0	0	0	0	0	0
DECL	0	10	3,33	30	10	3,33	0	0	0	3,33	0	33,33	3,33	0	0	0	0	3,33	0
DISA	0	0	0	6,67	80	0	0	0	0	0	0	0	0	0	13,33	0	0	0	0
DOUB	0	0	0	0	0	56,67	0	13,33	3,33	10	0	0	0	3,33	3,33	0	6,67	3,33	0
IRRI	0	36,67	0	0	0	0	60	0	0	0	0	3,33	0	0	0	0	0	0	0
IRON	0	0	0	6,67	23,33	10	0	13,33	6,67	23,33	0	0	0	0	10	0	3,33	3,33	0
INTI	20	0	0	16,67	0	0	0	0	26,67	0	0	6,67	13,33	6,67	0	6,67	0	0	3,33
CONT	0	3,33	0	6,67	10	16,67	0	10	0	20	0	3,33	0	10	6,67	0	10	3,33	0
IDS	0	0	0	0	0	0	0	0	0	3,33	10	3,33	3,33	0	0	76,67	0	3,33	0
OBVI	0	23,33	3,33	36,67	6,67	0	0	0	6,67	0	0	20	3,33	0	0	0	0	0	0
POLI	0	0	3,33	30	0	0	0	0	13,33	0	3,33	0	33,33	0	10	6,67	0	0	0
QUES	0	0	0	0	0	30	0	0	3,33	0	0	0	3,33	36,67	0	0	0	20	6,67
RESI	0	3,33	0	16,67	33,33	0	0	3,33	6,67	3,33	0	6,67	3,33	6,67	13,33	3,33	0	0	0
SEDU	6,67	0	0	10	0	0	0	0	10	0	10	0	26,67	0	0	36,67	0	0	0
NEG-S	0	0	0	0	0	33,33	3,33	0	0	0	0	0	0	0	0	0	13,33	33,33	16,67
NEU-S	3,33	0	0	0	0	20	0	0	0	3,33	0	13,33	3,33	13,33	0	0	6,67	23,33	13,33
POS-S	0	0	0	0	0	3,33	0	0	0	0	0	0	0	3,33	0	0	0	10	83,33

9syllabes	ADMI	AUTH	CONF	DECL	DISA	DOUB	IRRI	IRON	INTI	CONT	IDS	OBVI	POLI	QUES	RESI	SEDU	NEG S	NEU S	POS S
ADMI	0	0	6,67	16,67	3,33	3,33	0	0	10	0	0	13,33	16,67	3,33	0	0	3,33	20	3,33
AUTH	0	23,33	0	10	10	0	40	3,33	0	6,67	0	6,67	0	0	0	0	0	0	0
CONF	0	10	6,67	56,67	3,33	0	0	3,33	0	3,33	0	6,67	0	3,33	3,33	0	0	3,33	0
DECL	0	6,67	3,33	70	3,33	0	0	3,33	0	0	0	10	3,33	0	0	0	0	0	0
DISA	0	0	0	10	86,67	0	0	0	0	0	0	0	0	0	3,33	0	0	0	0
DOUB	0	0	3,33	10	0	33,33	0	13,33	0	3,33	0	0	3,33	6,67	0	0	10	16,67	0
IRRI	0	6,67	3,33	6,67	0	0	60	0	0	0	0	23,33	0	0	0	0	0	0	0
IRON	0	0	0	10	3,33	6,67	0	20	6,67	36,67	0	3,33	3,33	3,33	0	0	0	6,67	0
INTI	0	6,67	3,33	10	3,33	0	13,33	0	3,33	26,67	0	23,33	3,33	0	0	0	3,33	3,33	0
CONT	0	0	0	10	3,33	6,67	0	43,33	0	26,67	0	6,67	0	0	0	0	3,33	0	0
IDS	0	0	3,33	10	0	0	0	0	6,67	3,33	10	0	13,33	0	0	46,67	0	6,67	0
OBVI	0	16,67	3,33	20	20	0	20	3,33	3,33	6,67	0	3,33	0	0	3,33	0	0	0	0
POLI	0	0	0	50	0	0	0	0	6,67	0	10	6,67	26,67	0	0	0	0	0	0
QUES	0	0	0	3,33	0	23,33	0	10	0	0	0	0	0	56,67	0	0	0	6,67	0
RESI	0	0	0	6,67	50	0	0	3,33	6,67	0	0	3,33	0	0	26,67	0	3,33	0	0
SEDU	3,33	0	3,33	13,33	0	0	0	0	6,67	0	6,67	0	23,33	0	0	43,33	0	0	0
NEG-S	0	3,33	0	3,33	0	13,33	6,67	6,67	3,33	3,33	0	3,33	0	0	0	26,67	26,67	3,33	0
NEU-S	0	0	3,33	10	0	16,67	0	6,67	6,67	10	0	13,33	6,67	3,33	0	0	3,33	20	0
POS-S	3,33	0	10	3,33	0	10	0	3,33	6,67	0	0	10	3,33	0	0	3,33	3,33	26,67	16,67

Table 12 : Matrices de confusion des 19 affects sociaux à la longueur de 1, 2, 4 et 9 syllabes. Le taux de reconnaissance des affects sociaux est en gris foncé et les taux de confusion supérieurs à quatre fois le hasard (20%) sont en gris clair.

Dans un deuxième temps, quelques confusions importantes dont le score est supérieur à quatre fois le hasard, soit supérieur à 20%, sont remarquées et sont détaillées par la suite.

Certaines confusions sont trouvées pour toutes les longueurs d'énoncé: le report de la *confiance* (CONF), de l'*évidence* (OBVI) et de la *politesse* (POLI) vers la *déclaration* (DECL) ; la confusion entre l'*autorité* (AUTH) et l'*irritation* (IRRI), celle entre l'*ironie* (IRON) et le *mépris* (CONT), ainsi que celle entre la *résignation* (RESI) et la *déception* (DISA).

Les expressions de *confiance*, d'*évidence*, de *politesse* et de *déclaration*, montrent presque le même mouvement de contour de F0 pour toutes les longueurs, mais seules les confusions en direction de la *déclaration* sont importantes perceptivement, et ne sont pas dans le sens inverse, ni entre les autres trois attitudes. Cela confirme l'hypothèse proposée concernant l'attractivité perceptive de la *déclaration* pendant la présentation audio seule des affects sociaux chinois par des sujets natifs : les auditeurs ont à effectuer une tâche cognitive complexe (choisir une étiquette parmi 19). En face d'affects sociaux acoustiquement similaires, ils choisissent l'étiquette de *déclaration* comme choix par défaut. La figure 30 présente les comparaisons des courbes intonatives de la *confiance* avec celles de la *déclaration* pour chaque longueur et des exemples de la comparaison

des courbes intonatives entre la *confiance*, l'*évidence* et la *politesse* pour la phrase de neuf syllabes.

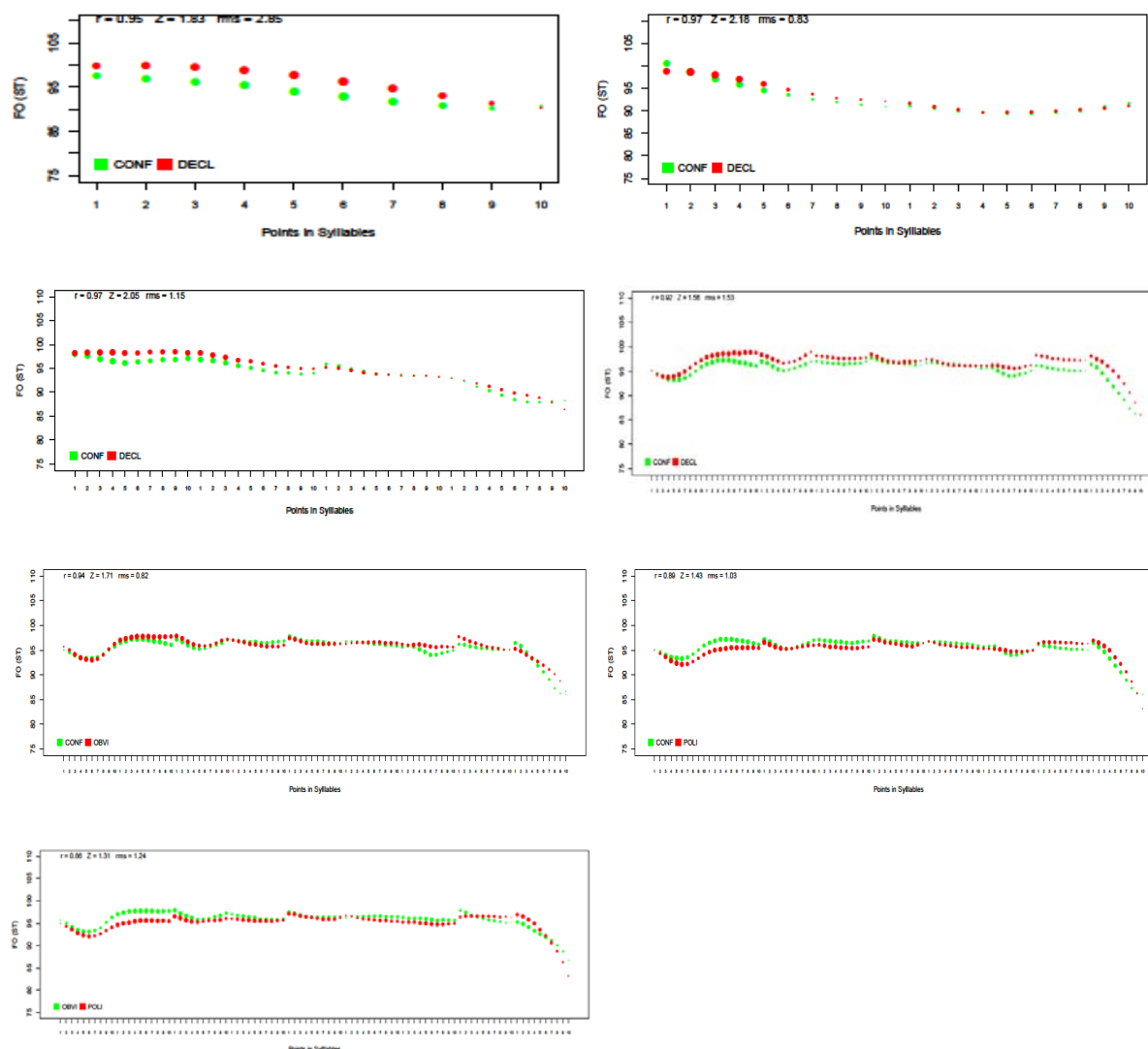


Figure 30 : Comparaison du contour intonatif de la *confiance* avec celui de la *déclaration* pour chaque longueur, et la comparaison des contours intonatifs entre la *confiance*, l'*évidence* et la *politesse* pour la phrase de 9 syllabes. Les valeurs de F0 des contours sont représentées en demi-ton et normalisés à 10 points par voyelle.  $r$  ( $-1 < r < 1$ ): coefficient de corrélation pondéré entre deux courbes intonatives;  $Z$ : score de la transformation de Fisher;  $rms$  ( $rms > 0$ ): différence moyenne quadratique entre deux courbes intonatives.

L'*autorité* et l'*irritation*, expressions largement confondues, présentent plus de similarités de forme de contour intonatif que de registre de F0. Mis à part la corrélation entre la forme de leurs courbes intonatives pour l'énoncé monosyllabique ( $r = 0,88$ ), celles obtenues pour les énoncés de deux, quatre et neuf syllabes sont toutes supérieures à 0,95 :  $r = 0,96, 0,99, 0,99$ . Par contre, l'expression d'*irritation* est marquée

par un registre plus élevé que celle d'*autorité*. La figure 31 présente la comparaison entre les contours intonatifs de l'*autorité* et de l'*irritation* pour chaque longueur.

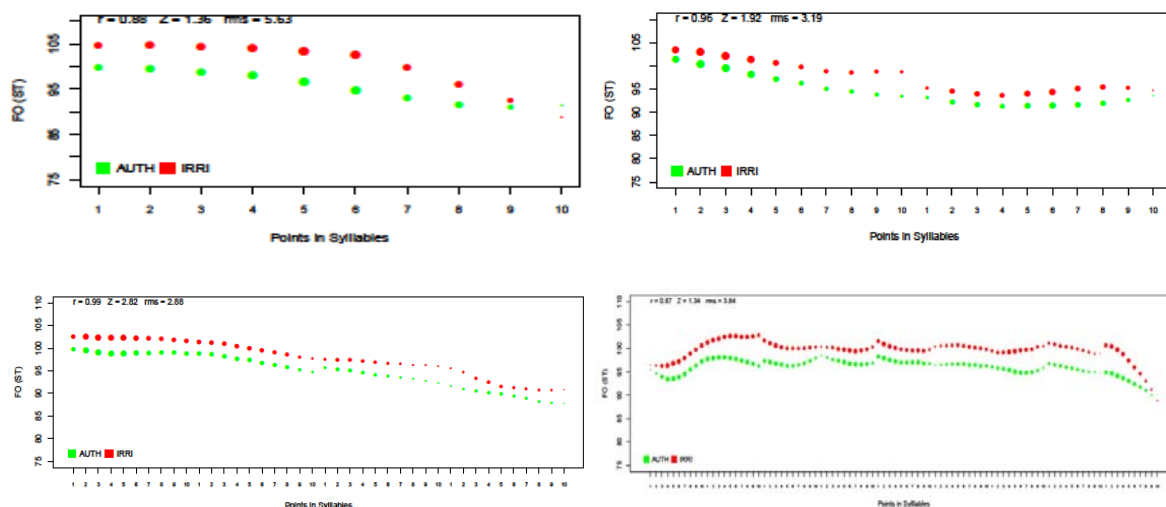


Figure 31 : Comparaison du contour intonatif de l'*autorité* avec celui de l'*irritation* pour chaque longueur. Les valeurs de F0 des contours sont représentées en demi-ton et normalisés à 10 points par voyelle.  $r$  ( $-1 < r < 1$ ): coefficient de corrélation pondéré entre deux courbes intonatives;  $Z$ : score de la transformation de Fisher ;  $rms$  ( $rms > 0$ ): différence moyenne quadratique entre deux courbes intonatives.

Les expressions de *déception* et de *résignation* sont très similaires au niveau du registre comme au niveau de la forme du contour intonatif. Cette observation confirme le résultat de l'analyse statistique des indices prosodiques des affects sociaux (voir supra).

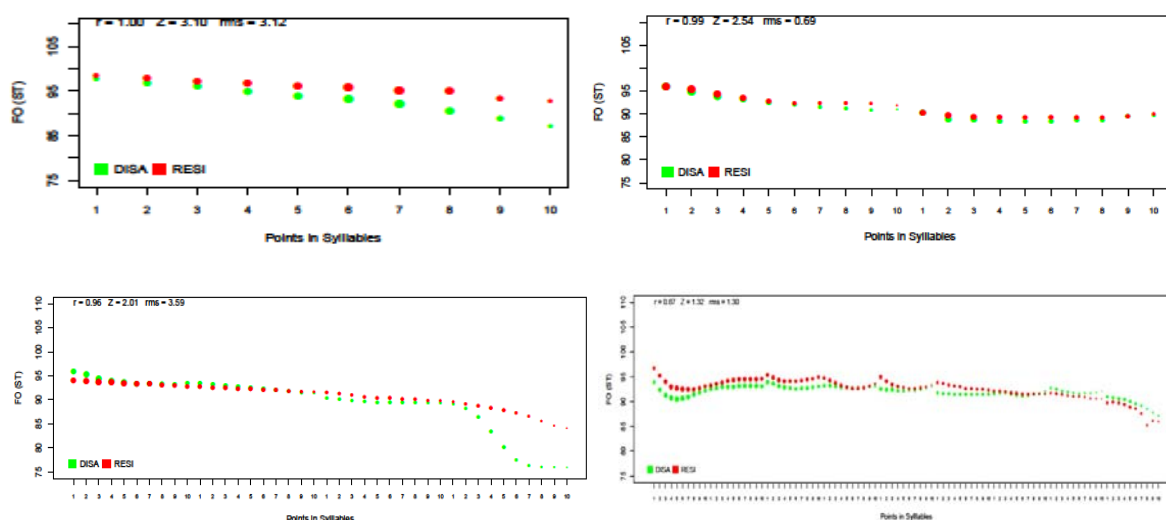


Figure 32 : Comparaison du contour intonatif de la *déception* avec celui de la *résignation* pour chaque longueur. Les valeurs de F0 des contours sont représentées en demi-ton et normalisés à 10 points par voyelle.  $r$  ( $-1 < r < 1$ ): coefficient de corrélation pondéré entre deux courbes intonatives;  $Z$ : score de la transformation de Fisher ;  $rms$  ( $rms > 0$ ): différence moyenne quadratique entre deux courbes intonatives.

Nous pouvons ainsi conclure que la confusion perceptive entre la *déception* et la *résignation* repose plutôt sur leur similarité des caractéristiques de F0 et du mouvement de contour intonatif. La figure 32 illustre les contours intonatifs des expressions de *déception* et de *résignation* pour chaque longueur.

La confusion entre l'*ironie* et le *mépris* est aussi observée pour toutes les quatre longueurs d'énoncé. Les contours de ces deux attitudes sont similaires pour les phrases de toutes longueurs, alors que le registre du *mépris* est plus élevé que l'*ironie*. Ceci pourrait constituer un indice expliquant leur confusion perceptive, mais il existe éventuellement autre paramètre prosodique comme la qualité de voix, qui pourrait aussi participer de cette confusion perçue. La figure 33 présente les contours intonatifs des expressions de l'*ironie* et du *mépris* pour chaque longueur.

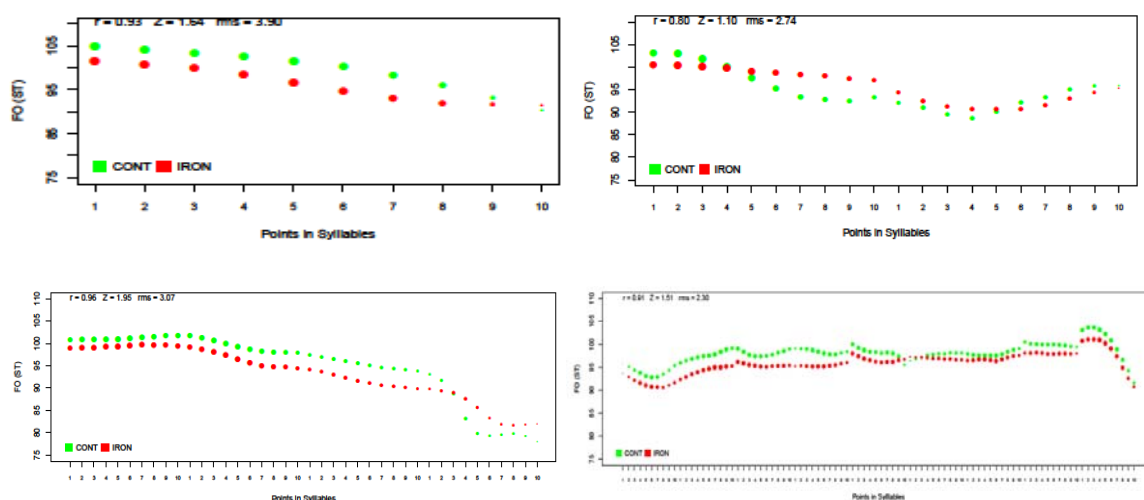


Figure 33 : Comparaison du contour intonatif de l'*ironie* avec celui du *mépris* pour chaque longueur. Les valeurs de F0 des contours sont représentées en demi-ton et normalisés à 10 points par voyelle.  $r$  ( $-1 < r < 1$ ): coefficient de corrélation pondéré entre deux courbes intonatives;  $Z$ : score de la transformation de Fisher;  $rms$  ( $rms > 0$ ): différence moyenne quadratique entre deux courbes intonatives.

Mis à part les confusions observées pour toutes les longueurs de phrase, il est intéressant de noter que l'expression de *question* est plus souvent confondue avec la *déclaration* lorsque l'énoncé est court (une et deux syllabes), alors qu'elle se fait des reports importants vers le *doute* lorsque l'énoncé est long (quatre et neuf syllabes). Il semble que, pour une expression interrogative, les énoncés courts sont moins à même de transmettre, via leurs caractéristiques prosodiques, cette sémantique interrogative lorsque le contexte de communication est omis.

#### 2.4.6. Conclusion

Dans cette section, nous avons d'abord examiné les principaux traits prosodiques des 19 affects sociaux chinois via une analyse des trois paramètres acoustiques classiques que sont la fréquence fondamentale, l'intensité et la durée. Nous avons ensuite complété notre étude acoustique en observant les contours intonatifs des affects sociaux. 76 énoncés qui avaient été validés et avaient obtenus les meilleurs scores de reconnaissance à la perception de 30 auditeurs natifs ont été retenus pour cette analyse. Les principaux résultats trouvés sont les suivants :

En premier lieu, et selon le résultat du classement des affects sociaux obtenu à partir d'une classification hiérarchique basée sur leur dispersion obtenue par une ACP, les 19 affects sociaux sont organisés en deux groupes principaux. Cette séparation est fortement liée à leurs caractéristiques de fréquence fondamentale : un des groupes est composé des affects sociaux caractérisés par un registre de F0 élevé (par exemple l'exclamation de *surprise positive* et l'*admiration*), par une plage de F0 étendue (par exemple l'exclamation de *surprise négative* et le *doute*) et par une importante variation de F0 (par exemple l'exclamation de *surprise neutre* et *mépris*). L'autre groupe contient les affects sociaux dont le registre, la plage de F0 et la variation de F0 sont, respectivement plus bas, plus étroit et moins important (par exemple la *déception*, la *résignation*, la *déclaration*, l'*autorité* et la *politesse*). En général, cette observation confirme l'hypothèse d'une corrélation entre la fréquence fondamentale et le message affectif transmis par la prosodie (Bolinger, 1964, 1978; Ohala, 1983). Selon cette hypothèse, les fréquences fondamentales basses sont liées à des expressions de dominance ou d'assertivité, de confiance, d'autorité, d'agression et de protection, tandis que les fréquences fondamentales hautes sont liées plutôt à des expressions de déférence, d'amicalité, de manque de confiance, d'insécurité et de vulnérabilité. D'ailleurs, cela rejoint aussi les propositions de Caelen-Haumont (1991) et de Kehrein (2002) : plus la F0 est élevée et plus la force du caractère inattendu de l'expression est importante.

Un constat intéressant différent doit être fait pour le regroupement des affects de soumission (la *déception* et la *résignation*) avec l'*autorité* et la *déclaration*. Ces expressions forment une catégorie d'affects sociaux avec la plage de F0 la plus réduite et

le registre le plus bas. Ces observations ne sont pas explicables par les prédictions du « Code de Fréquence » de Ohala (1983), qui propose que les affects de soumission soient produits avec un registre de F0 plutôt élevé. Dans ce cas-là, l'interprétation proposée par le « Code d'Effort » nous semble une explication possible pour comprendre les observations prosodiques de ces deux attitudes. Selon Gussenhoven (2004), les affects de soumission seront produit à l'inverse des expressions autoritaires, avec une F0 basse et peu dynamique. Une étude plus poussée de ces phénomènes est nécessaire pour permettre des explications plus détaillées et constitue l'une de nos pistes de recherche future.

Lors de l'analyse acoustique des contours intonatifs de ces affects sociaux, nous constatons également une distinction entre deux groupes d'affects sociaux quasi identiques pour tous les énoncés en fonction du registre de F0 comme en fonction de la forme de contour. Une majorité des affects sociaux est classée dans le même groupe, indépendamment des variations de longueur des énoncés, à la fois au niveau de RMS et de la corrélation de forme de contour. Il s'agit des expressions de *déclaration*, d'*évidence*, d'*autorité*, de *politesse*, de *confiance*, de *résignation* et de *déception* d'un côté, ainsi que des expressions de *surprise-négative*, de *surprise-neutre*, de *surprise-positive*, d'*admiration* et de *question* de l'autre côté.

Une autre observation intéressante porte sur la forme globale des contours intonatifs des affects sociaux : plus l'énoncé est long, plus les contours intonatifs se différencient. D'ailleurs, pour les affects qui sont proches en registre mais se différencient au niveau de leur forme, la différence principale consiste essentiellement en début et en fin de phrase. Ceci confirme l'observation de Mac et al. (2012) sur les contours locaux et globaux des affects sociaux vietnamiens et conforte également notre observation des caractéristiques prosodiques de la première et de la dernière syllabe dans l'analyse des indices prosodiques des affects sociaux chinois.

Une séparation similaire entre ces mêmes affects sociaux a également été constatée sur la base des données de perception par des chinois natifs. Ceci indique le rôle majeur de F0 dans le décodage des expressions affectives. De plus, si l'on compare les catégories perceptives et les catégories acoustiques des mêmes affects sociaux, on découvrira des liens qui nous permettent d'expliquer des confusions perceptives entre les affects



sociaux. Par exemple, les confusions entre la *déclaration* et la *confiance*, la *politesse* et l'*évidence* pourraient avoir comme origine leur homogénéité en fréquence fondamentale et leur similarité en terme de mouvement intonatif ; ceci valant aussi pour expliquer la confusion entre la *déception* et la *résignation*. La similarité entre la forme des contours intonatifs de l'*autorité* et de l'*irritation* pourrait aussi fournir une explication à leur confusion perceptive. Pourtant, il est difficile d'expliquer la confusion entre le *maternel* et la *séduction* par leurs corrélations acoustiques et leurs mouvements intonatifs. C'est en l'occurrence la qualité de voix qui pourrait nous donner une piste explicative.

## **2.5. Perception multimodale des affects sociaux chinois**

### **2.5.1. Introduction**

Dans la littérature des recherches sur les expressions multimodales des émotions (par exemple Scherer et al., 2003 ; Scherer & Ellgring, 2007), des affects sociaux (par exemple Granström & House, 2005), et de la multimodalité de la prosodie (par exemple Swerts & Krahmer, 2005 ; Barkhuysen et al., 2007), il a été proposé que les émotions et les attitudes sont véhiculées à l'aide de la prosodie dans une communication face-à-face, et la prosodie est souvent produite et perçue par le biais des signaux acoustiques et des expressions faciales.

Inspirés par les recherches antérieures, nous voulons continuer notre recherche en nous intéressant à la perception multimodale des affects sociaux chinois, dans le but de vérifier le rôle et la contribution relative de l'audition et de la vision pour la perception des affects sociaux dans une communication face-à-face. Ceci est important pour l'apprentissage des langues étrangères, car les apprenants ont toujours besoin de se situer dans une situation face-à-face, dynamique et interactive. Pourtant, considérant la complexité de la tâche que le sujet effectue lors d'un test perceptif multimodal, et en vue de la faciliter, nous réduisons le nombre d'affects sociaux dans le présent test perceptif en choisissant ceux qui sont les plus représentatifs et qui pourraient être les plus problématiques pour les apprenants français du chinois langue étrangère. Les affects sociaux sélectionnés sont identiques à ceux utilisés dans le test multimodal des expressions socio-affectives en français par des apprenants chinois, que nous présenterons dans le chapitre 4.

## 2.5.2. Modalité de l'expression des affects sociaux

L'objectif principal de cette étude vise à évaluer l'influence relative des deux modalités auditive et visuelle et de ces deux modalités combinées sur la perception des affects sociaux. Un test perceptif est donc réalisé, qui propose trois conditions différentes de présentation des affects sociaux : « audio seul », « vidéo seul » et « audio-vidéo ».

Dans la condition « audio seul » les sujets doivent juger l'affect social exprimé à l'aide des seules informations transmises via la voix de la locutrice ; dans la condition « vidéo seul », ils doivent reconnaître l'affect social exprimé à partir des seules informations véhiculées par l'expression faciale et les mouvements du buste de la locutrice ; dans la condition « audio-vidéo », les sujets forment leurs jugements à l'aide de ces deux informations combinées.

## 2.5.3. Sous-corpus

### 2.5.3.1. Sélection des affects sociaux

Lors des expériences de perception auditive des 19 affects sociaux chinois effectuées auprès de sujets natifs (voir partie 2.3) et non natifs (que nous présenterons dans le chapitre 3) sur des stimuli audio, nous avons constaté que généralement le taux d'identification des affects sociaux n'était pas très élevé. Ce résultat est probablement lié à la tâche cognitivement complexe proposée aux sujets : faire un choix parmi 19 étiquettes possibles. En vue d'alléger cette charge cognitive et de diminuer le bruit induit, qui peut biaiser les analyses, nous réduisons le nombre d'affects sociaux en nous concentrant sur un sous-ensemble composé de ceux qui attirent plus de confusions perceptives et qui pourraient induire également des différences entre les sujets chinois et français. La sélection se base sur le taux d'identification et de l'attractivité des affects sociaux, tels qu'obtenus lors du test de reconnaissance.

Dans un premier temps, l'affect social qui n'a pas été reconnu (dont le taux de reconnaissance est inférieur au seuil du hasard) ni par les sujets chinois ni par les sujets français est enlevé. Il s'agit de l'attitude de *confiance* (CONF) (voir figure 34).

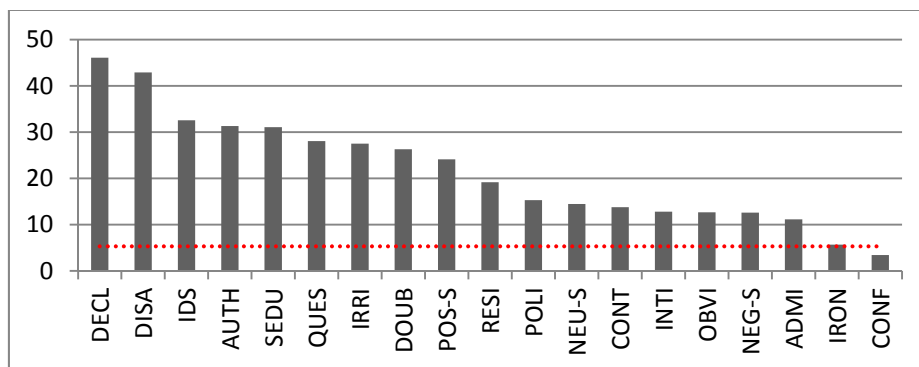


Figure 34: Taux de reconnaissance des 19 affects sociaux chinois pour les sujets chinois et français. La ligne pointillée rouge représente le seuil du hasard (5,3%).

Dans un deuxième temps, les affects sociaux qui ont attiré le plus de réponses erronées durant les tests perceptifs sont sélectionnés. La figure 35 montre le taux d'attractivité des affects sociaux pour les sujets chinois et français. La *déclaration* (DECL) est l'attitude la plus « attractive » parmi les 19, et il existe un grand décalage entre la *déclaration* et les autres affects. Dans ce cas, nous considérons pertinent de calculer le taux d'attractivité moyen des 18 affects sociaux et de choisir ceux dont le taux d'attractivité est supérieur à la moyenne. Par conséquent, à part la *déclaration*, huit attitudes sont choisies pour le présent test multimodal en suivant ce critère: l'*évidence* (OBVI), le *doute* (DOUB), la *question* (QUES), l'*autorité* (AUTH), l'exclamation de *surprise neutre* (NEU-S), la *déception* (DISA), l'*intimité* (INTI) et le *mépris* (CONT).

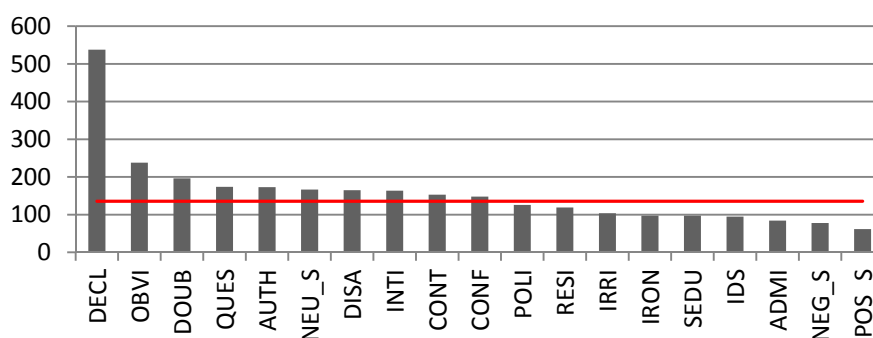


Figure 35: Taux d'attractivité des 19 affects sociaux chinois pour les sujets chinois et français. La ligne pleine rouge représente le taux moyen d'attractivité calculé avec les 18 affects sociaux sauf la *déclaration*.

Dans un troisième temps, tenant compte que les affects sociaux sélectionnés seront également utilisés lors d'expériences interculturelles dédiées à la recherche sur l'apprentissage de la prosodie socio-affective en français et chinois langue étrangère ou seconde, nous considérons convenable de sélectionner les affects sociaux dont le taux de

reconnaissance montre un grand décalage entre les sujets chinois natifs et les sujets français. En conséquence, l'*irritation* (IRRI), expression pour laquelle la différence de taux de reconnaissance entre les deux groupes de sujets est importante, a été ajoutée à la liste d'affects sociaux sélectionnés (figure 36).

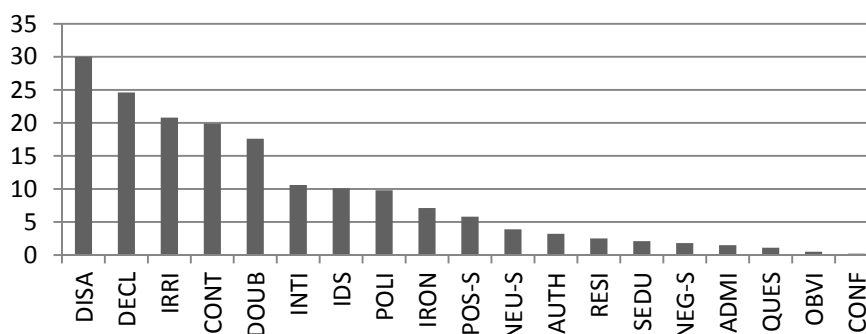


Figure 36: La différence de taux d'identification des 19 affects sociaux chinois entre les sujets chinois et français.

Puisque que la *politesse* est un affect typique de la relation sociale entre le locuteur et interlocuteur lors de l'interaction, et qu'elle joue un rôle important dans la vie sociale, nous l'ajoutons aussi pour représenter, avec l'*autorité*, les affects qui véhiculent des valeurs sociales. Finalement, parmi les affects sociaux étroitement liés au contexte social spécifique dans lequel l'interaction se déroule, le *maternel* est plus précisément défini par le contexte que l'*intimité* et la *séduction*, il est aussi jugé de meilleure qualité par la locutrice elle-même. Par conséquent, nous avons décidé de remplacer l'*intimité* par le *maternel*.

Conformément aux critères détaillés ci-dessus, onze affects sociaux sont sélectionnés pour les expériences multimodales croisant le chinois mandarin et le français. Ils sont la *déclaration* (DECL), l'*évidence* (OBVI), le *doute* (DOUB), la *question* (QUES), l'*autorité* (AUTH), l'exclamation de *surprise neutre* (NEU-S), la *déception* (DISA), le *mépris* (CONT), l'*irritation* (IRRI), la *politesse* (POLI) et le *maternel* (IDS).

### 2.5.3.2. Composition du sous-corpus

Un seul énoncé a été retenu pour le test perceptif audiovisuel sur le seul critère du résultat de perception pour l'ensemble des affects sociaux pour les sujets français, que nous allons présenter dans le chapitre 3.

Lors de l'expérience auditive des affects sociaux chinois menée avec des sujets natifs, les énoncés de quatre syllabes ont reçu les meilleurs scores d'identification, - résultat confirmé avec les sujets français et vietnamiens. Parmi les 12 énoncés de quatre syllabes que les sujets français ont écoutés, celui de séquence tonale 4-1-1-4 a obtenu de bons scores moyens pour tous les affects sociaux et à la fois presque aucun mauvais score de reconnaissance. En ce sens, cet énoncé peut être considéré comme celui véhiculant les affects sociaux les mieux reconnus parmi les énoncés du corpus. Les expressions socio-affectives réalisées avec cet énoncé ont donc été sélectionnées pour former le sous corpus audiovisuel. D'un point de vue de sa structure tonale, cet énoncé porte un ton initial et final descendant (ton 4), tandis que toutes les autres syllabes portent un ton plat (ton 1). D'un point de vue syntaxique, il est composé d'un groupe nominal de quatre syllabes. L'énoncé sélectionné est présenté ci-dessous avec sa transcription phonétique en pinyin et sa traduction française :

« 四天三夜 », si4 tian1 san1 ye4, « quatre jours trois nuits ».

Si nous avons choisi de prendre une seule phrase pour le présent test perceptif, c'est parce que nous cherchons ici à examiner principalement l'effet de la modalité sur les expressions des affects sociaux, et que les facteurs linguistiques ne sont pas pris en considération dans la présente étude. Par conséquent, le sous-corpus audiovisuel est finalement composé de 33 stimuli : un énoncé de quatre syllabes réalisé avec les onze affects sociaux présentés selon les trois modalités « audio seul », « vidéo seul » et « audio-vidéo ».

#### 2.5.4. Sujets

32 participants chinois natifs (16 hommes et 16 femmes), d'âge moyen de 33 ans, ont participé à cette expérience. Tous les sujets sont professeurs à l'Université des Etudes internationales de Xi'an, à Xi'an en Chine. Aucun n'a signalé de trouble auditif ou visuel.

Durant le test, les sujets ont été aléatoirement divisés en deux groupes, équilibrés en genre. Les modalités sont présentées à chacun des groupes selon un ordre différent: le premier groupe (8 hommes et 8 femmes) a passé les trois modalités dans l'ordre « audio seul », « vidéo seul », puis « audio-vidéo » ; le second groupe (8 hommes et 8 femmes) a passé les trois modalités dans l'ordre « vidéo seul », « audio seul », puis « audio-vidéo ».

### 2.5.5. Protocole expérimental

Le test perceptif s'est déroulé via une interface informatique, réalisée à l'aide du logiciel « LiveCode », pour Windows. Le test débutait par une présentation globale et une description de chaque affect social illustré par des exemples de contextes possibles. Ces informations étaient écrites dans la langue maternelle du sujet. Les sujets conservaient ce document comme référence pendant le déroulement de test. Après avoir lu les consignes et avoir effectué un entraînement, les sujets commençaient le test. Les affects sociaux ont été ensuite présentés dans chacune des trois modalités de présentation, modalités elles-mêmes présentées dans un ordre donné pour chaque groupe (voir supra). A l'intérieur de chaque condition expérimentale, chaque affect social a été présenté une seule fois, dans un ordre aléatoire, différent pour chaque sujet. Pour chaque stimulus, les sujets devaient juger quel est affect social exprimé par la locutrice en choisissant une étiquette parmi les onze proposés à l'écran. Aucune limitation de temps de réflexion avant la présentation du stimulus suivant n'a été imposée. Les réponses données par les sujets ont été sauvegardées automatiquement sur le disque dur de l'ordinateur.

### 2.5.6. Analyse des résultats

#### 2.5.6.1. Facteurs influant la reconnaissance des affects sociaux

Dans un premier temps, les réponses de sujets ont été codées avec deux chiffres : 0 représente une mauvaise réponse (le participant ne reconnaît pas l'affect social présenté) ; 1 représente une bonne réponse (le participant reconnaît l'affect social présenté). Une régression logistique a été effectuée sur ces données pour étudier l'effet de plusieurs facteurs sur cette variable réponse binaire. Les facteurs fixés sont la modalité de l'expression socio-affective (« audio seul », « vidéo seul » et « audio-vidéo »), l'ordre de présentation des modalités (« audio seul », « vidéo seul », puis « audio-vidéo » ou « vidéo seul », « audio seul », puis « audio-vidéo »), le genre du sujet (homme ou femme) et l'affect social présenté (11 affects sociaux). Le modèle de régression logistique est une approche standard pour analyser la relation entre une variable dépendante binaire et des variables explicatives, dont le grand intérêt est d'évaluer l'impact des variables explicatives sur la probabilité du risque lié à la variable dépendante (Lu & Yang, 2012).

Cependant, dans notre expérience un participant est interrogé plusieurs fois sur une même valeur d'affect social. Nous recueillons donc plusieurs valeurs de la variable réponse pour un même sujet. Cette situation ne garantit pas l'indépendance des observations et certaines différences sur la variable réponse peuvent donc avoir pour origine cette dépendance. Dès lors, comme nous sommes en présence d'une variable réponse binaire et de mesures répétées, nous avons choisi d'utiliser une régression logistique mixte. C'est une extension de la régression logistique pour laquelle des effets aléatoires se rajoutent aux effets fixes. Les effets fixes sont les effets des facteurs détaillés ci-dessus et sont communs à l'ensemble des individus. Les effets aléatoires représentent quant à eux les effets des individus. Cette méthode permet alors d'être plus précis concernant l'explication des variations de la variable réponse. La régression logistique mixte a été réalisée à l'aide de la fonction « glmer » du package « lme4 » du logiciel R.

Une procédure descendante (*backward* en anglais) a été utilisée pour sélectionner les variables significatives. Le modèle initial de la procédure comprend les quatre facteurs et leurs interactions d'ordre 2. A chaque étape, la variable dont la valeur de  $p$  de l'effet associée au test de rapport de vraisemblance est la plus grande et supérieure à 0,05 est retirée du modèle. Le processus s'arrête lorsque toutes les variables sont retirées du modèle ou lorsque la valeur de  $p$  est inférieure à 0,05. Lorsqu'un facteur est impliqué dans une interaction significative, il ne peut pas être retiré du modèle même s'il est testé comme non significatif. Cette méthode sera réutilisée pour exploiter les données obtenues dans les autres expériences de perception multimodale des affects sociaux, que nous allons présenter dans le chapitre 4.

La table 13 donne le résultat de la sélection descendante. La première colonne présente le nom de l'effet qui est retiré à chaque étape. Les trois colonnes suivantes donnent respectivement la valeur de la statistique du test, le nombre de degrés de liberté et la probabilité du test de rapport de vraisemblance. Selon le tableau, la valeur de  $p$  des effets présentés est supérieure à la probabilité, ils sont donc tous retirés du modèle. Autrement dit, les effets restants ont un effet significatif sur la variable réponse.

	LRT $\chi^2$	ddl	<i>p</i>
Genre: Modalité	1,5284	2	0,4657
Genre : Ordre de présentation des modalités	0,6902	1	0,4061
Modalité : Ordre de présentation des modalités	2,1169	2	0,347
Genre : Affect social	12,177	10	0,2734
Genre	2,9989	1	0,08332

Table 13: Résultat de l'analyse de régression logistique mixte pas à pas descendante : les effets présentés dans le tableau sont testés comme **non significatifs** selon un test de rapport de vraisemblance ( $p > 0,05$ ).

Ainsi, les facteurs significatifs sont l'affect social, la modalité, l'ordre de présentation des modalités, ainsi que les interactions entre affect social et ordre de présentation des modalités, et entre l'affect social et modalité. Le genre du sujet n'a pas d'effet significatif sur les réponses des sujets.

#### 2.5.6.2. Effet de la modalité et de l'ordre de présentation des modalités

Suite à l'analyse de régression logistique mixte, nous cherchons ici à mettre en évidence, pour chaque affect social, les différences de proportions de bonne réponse entre les trois modalités et les deux ordres de présentation des modalités. Pour ce faire, nous avons appliqué la méthode de comparaison multiple, présentée en détail dans Hothorn et al. (2008). L'analyse a été réalisée à l'aide de la fonction « lsmeans » du package « lsmeans » du logiciel R.

- Effet de la modalité

Selon le résultat de la comparaison multiple, l'*autorité*, la *déception*, le *maternel*, la *politesse*, l'exclamation de *surprise neutre* et la *question* montrent des différences significatives entre les modalités en matière de reconnaissance. Pour l'*autorité*, le taux d'identification en « audio-vidéo » est significativement plus élevé qu'en « vidéo seul » ( $p < 0,05$ ) ; pour la *déception*, la progression du taux de reconnaissance est significative entre « audio seul » et « audio-vidéo » ( $p < 0,01$ ), entre « audio seul » et « vidéo seul » ( $p < 0,05$ ), ainsi qu'entre « vidéo seul » et « audio-vidéo » ( $p < 0,05$ ) ; pour le *maternel*, le taux est significativement plus élevé en audiovisuel qu'en vidéo seul ( $p < 0,05$ ) ; pour l'exclamation de *surprise neutre*, les taux de reconnaissance en « audio seul » et en « audio-vidéo » sont significativement supérieurs à celui obtenu en condition visuelle seule ( $p < 0,01$ ) ; pour la *politesse*, les taux de reconnaissance en modalité audiovisuelle



et visuelle seule sont significativement supérieurs à celui obtenu en audio seul ( $p < 0,05$ ) ; la *question* montre des taux significativement plus élevés en audio seul qu'en vidéo seul ( $p < 0,01$ ).

Ce résultat est illustré par la figure 37, qui représente le taux de reconnaissance moyen obtenu pour chacun des affects sociaux dans chacune des trois modalités.

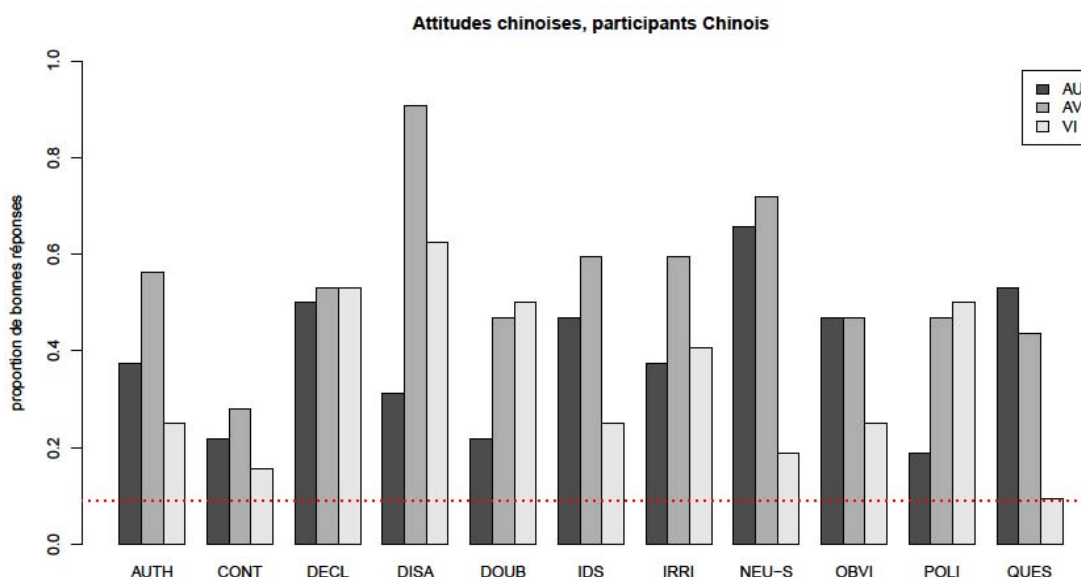


Figure 37: Effet de la modalité sur les taux d'identification des 11 affects sociaux chinois pour 32 sujets chinois. Les affects sociaux sont présentés en abscisse et les scores de reconnaissance en ordonnée. La ligne rouge pointillée indique le seuil du hasard (9%).

Généralement, tous les affects sociaux sont reconnus par les sujets natifs dans toutes les modalités (le taux d'identification est supérieur au seuil du hasard). Les résultats obtenus en modalité audiovisuelle montrent des meilleurs taux d'identification pour tous les affects sociaux sauf pour le *doute* (DOUB), la *politesse* (POLI) et la *question* (QUES). Par rapport à l'information auditive, l'information visuelle joue un rôle prédominant dans la reconnaissance de l'expression de *politesse*. A l'inverse, pour l'expression d'exclamation de *surprise neutre* (NEU-S) et de *question* (QUES), les sujets ont plus recours à l'information auditive qu'à l'information visuelle. Le cas de la *déclaration* (DECL) est un peu particulier car les taux de reconnaissance en trois modalités sont presque identiques, ce qui signifie que les informations auditive et visuelle sont aussi importantes pour l'expression de *déclaration*. Pour le reste des affects sociaux : le *mépris* (CONT), la *déception* (DISA), le *doute* (DOUB), l'*évidence* (OBVI) et

l'*irritation* (IRRI), l'information auditive et visuelle coopèrent et se complètent l'une avec l'autre dans l'expression. Parmi ces affects sociaux, la *déception* montre le plus de synergie audiovisuelle avec le taux d'identification le plus élevé parmi les onze affects sociaux.

- Effet de l'ordre de présentation des modalités

Selon le résultat de la comparaison multiple, l'ordre de présentation des modalités exerce un effet significatif sur la reconnaissance du *maternel* ( $p < 0,05$ ) et de l'exclamation de *surprise neutre* ( $p < 0,01$ ).

Comme le montre la figure 38, les deux groupes d'ordre de présentation des modalités ont des scores de reconnaissance similaires pour presque tous les affects sociaux. Le *maternel* (IDS) et l'exclamation de *surprise neutre* (NEU-S) montrent des différences entre les deux groupes de sujets, et les sujets qui ont passé en premier la modalité visuelle seule reçoivent des meilleurs scores de reconnaissance que les autres sujets qui ont passé d'abord la modalité audio seule.

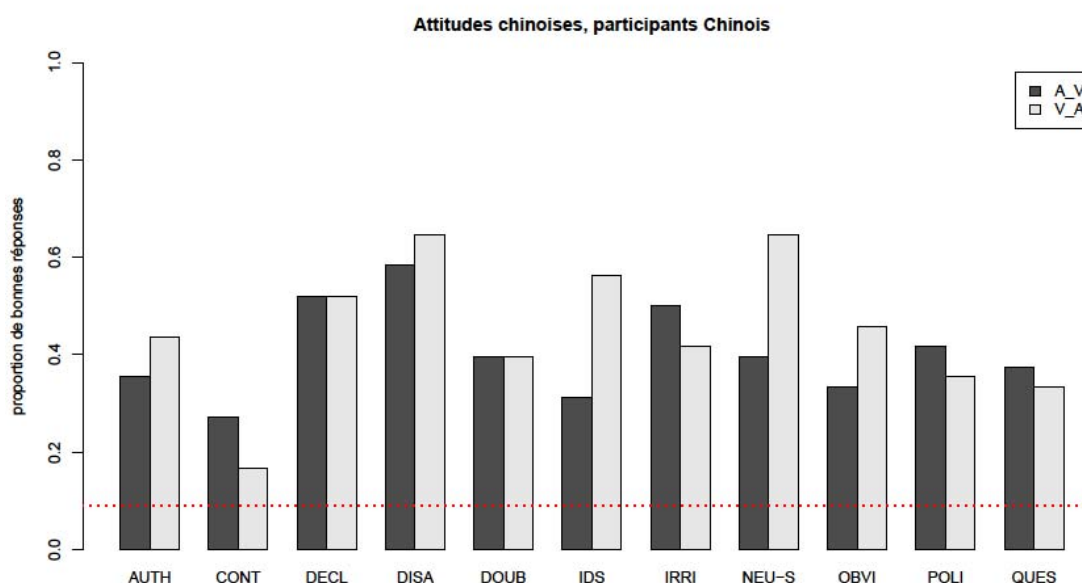


Figure 38: Effet de l'ordre de présentation des modalités sur les taux d'identification des 11 affects sociaux chinois pour 32 sujets chinois. Les affects sociaux sont présentés en abscisse et les scores de reconnaissance en ordonnée. La ligne rouge pointillée indique le seuil du hasard (9%).

### 2.5.6.3. Confusion perçue des affects sociaux

L'analyse des confusions entre les affects sociaux a été réalisée pour les 32 sujets chinois, sur la base de la matrice de confusions de chaque modalité. Les résultats des deux groupes de sujets ont été mélangés. Cette analyse a été effectuée en deux étapes : la première étape consiste à discuter les confusions les plus importantes entre affects sociaux et l'orientation de ces confusions, ce qui correspond bien à la problématique de la perception de ces affects sociaux dans une modalité donnée ; la seconde étape est basée sur une analyse de classification hiérarchique de ces expressions socio-affectives dont le but est d'extraire les tendances générales du regroupement des affects sociaux en considération de l'ensemble de confusions perceptives apparues lors de l'expérience.

Lors de l'analyse de classification hiérarchique, le carré de la distance Euclidienne a été utilisé comme distance entre les observations, qui ont été alors regroupées en classes différentes selon le critère de variance minimale de Wald. En vue de déterminer le nombre optimal de groupes d'affects sociaux sur l'arbre hiérarchique, nous nous sommes référés au critère de réduction de la variance (*elbow criterion* en anglais) expliquée par chaque groupe (Soni Madhulatha, 2012). Selon ces résultats, les onze affects sociaux peuvent être répartis en groupes qui varient selon la modalité considérée.

- Modalité « audio seul »

La figure 39 illustre les principales confusions perceptives entre les affects sociaux pour les 32 sujets dans la condition audio seule, obtenues à partir de la matrice de confusion. Les confusions deux fois supérieures au hasard (9%) sont proposées.

Selon le résultat, la *déception* (DISA), l'*évidence* (OBVI) et la *politesse* (POLI) sont principalement confondues avec la *déclaration* (DECL), qui montre quelques confusions avec l'*évidence* (OBVI). La *politesse* est confondue avec la *déclaration* avec un pourcentage important (56,3%). L'*irritation* (IRRI) n'est confondue qu'avec l'*autorité* (AUTH) qui se reporte sur l'*évidence*. Le *doute* (DOUB) et la *question* (QUES) sont confondus, tandis que le *mépris* (CONT) n'est particulièrement confondu qu'avec le *doute*. Le *maternel* (IDS) et l'exclamation de *surprise neutre* (NEU-S) sont bien reconnus et ne montrent pas de confusions particulières avec les autres affects sociaux.

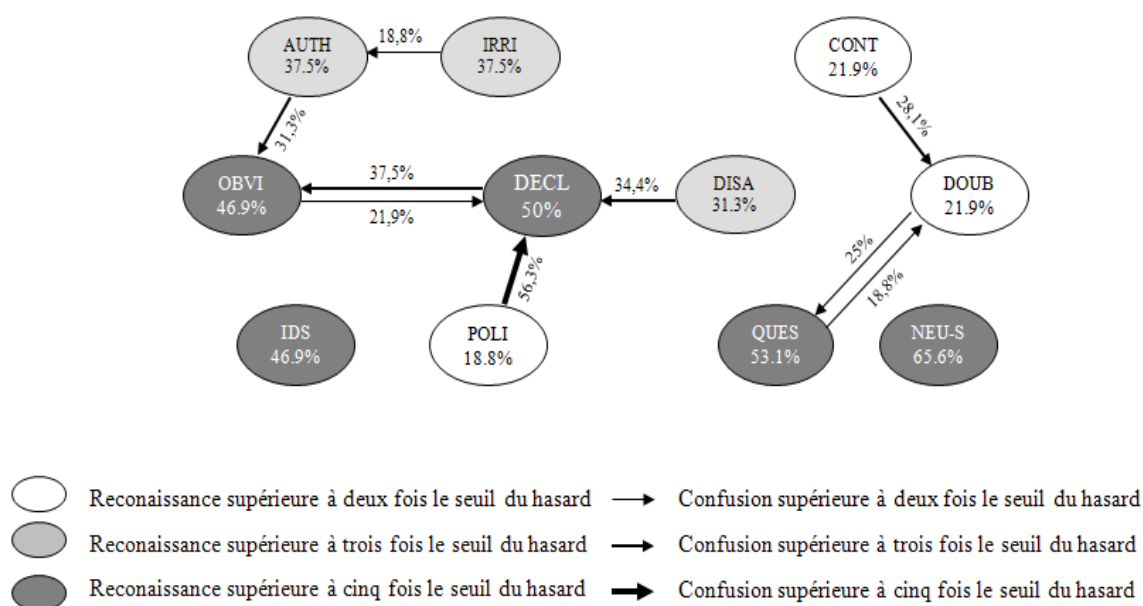


Figure 39: Confusions des 11 affects sociaux chinois pour 32 sujets natifs dans la modalité « audio seul ». Les pourcentages d'identification sont indiqués dans les ovales et les taux de confusions sur les flèches. Seules les confusions qui affichent des scores supérieurs à deux fois le seuil du hasard sont rapportées.

En considérant toutes les confusions entre affects sociaux, nous avons mené une analyse de classification hiérarchique afin d'identifier des classes perceptives de ces affects sociaux pour les sujets natifs dans la condition « audio seul ».

Le dendrogramme de la figure 40 représente les résultats de cette analyse. Le graphique situé en haut à droite du dendrogramme illustre la perte d'inertie en fonction du nombre de groupes, à partir duquel on peut conclure qu'il est raisonnable de considérer 5 catégories d'affects sociaux.

Selon cette analyse, la *déclaration* (DECL), la *politesse* (POLI) et la *déception* (DISA) sont regroupées dans la même catégorie ; l'*autorité* (AUTH) et l'*irritation* (IRRI), deux affects sociaux d'imposition sont regroupées avec l'*évidence* (OBVI) dans la deuxième catégorie ; ces deux groupes se relient à un niveau plus élevé éventuellement à cause des confusions réciproques entre la *déclaration* et l'*évidence* (voir figure 39). La troisième catégorie contient le *mépris* (CONT), le *doute* (DOUB) et la *question* (QUES). Le *maternel* (IDS) et l'exclamation de *surprise neutre* (NEU-S), affects sociaux bien reconnus par les sujets, forment des isolats faute de confusions particulières avec d'autres affects sociaux.

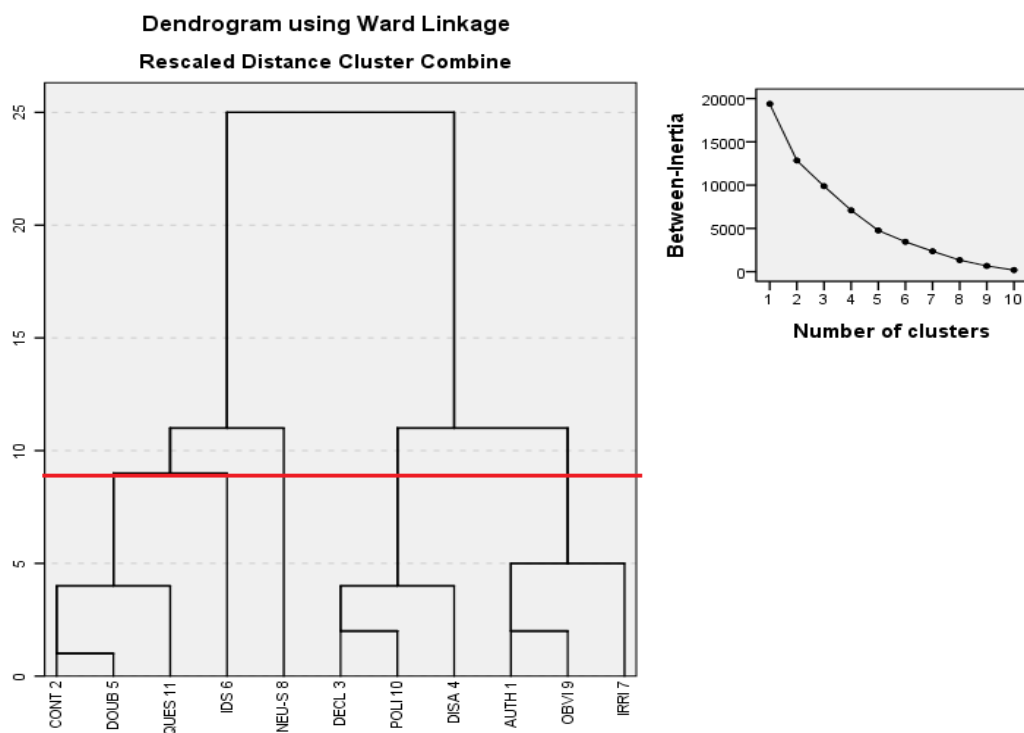


Figure 40: Classification hiérarchique des 11 affects sociaux chinois pour 32 sujets natifs dans la modalité « audio seul ».

- Modalité « vidéo seul »

Les principales confusions entre les affects sociaux pour tous les sujets dans la modalité « vidéo seul » sont schématisées sur la figure 41. La *déclaration* (DECL), la *politesse* (POLI), la *déception* (DISA) et le *doute* (DOUB) sont bien perçus, alors que les indices visuels ne permettent pas une reconnaissance efficace, dans notre cas, de la *question* (QUES) et du *mépris* (CONT). Les autres affects sociaux sont correctement identifiés.

Un taux de confusion plus important qu'en condition audio seule est observé dans cette condition expérimentale. L'*autorité* (AUTH) est confondue avec la *déclaration* et l'*évidence* (OBVI); la *politesse* est mélangée avec la *déclaration*, elle-même étant confondue avec l'*évidence*; le *maternel* (IDS) n'est confondu qu'avec la *politesse*. Le *doute* et l'exclamation de *surprise neutre* (NEU-S) montrent des reports vers la *question* et l'exclamation de *surprise neutre* est aussi confondue avec le *doute*. On observe aussi des reports de l'*irritation* (IRRI) et du *mépris* vers la *déception* qui est confondue avec le *mépris*, ainsi que l'*irritation*. Il est intéressant de remarquer que la *question* n'est pas bien reconnue dans la modalité visuelle seule (9,4%) et entretient d'importantes confusions avec la *déclaration* (43,8%).

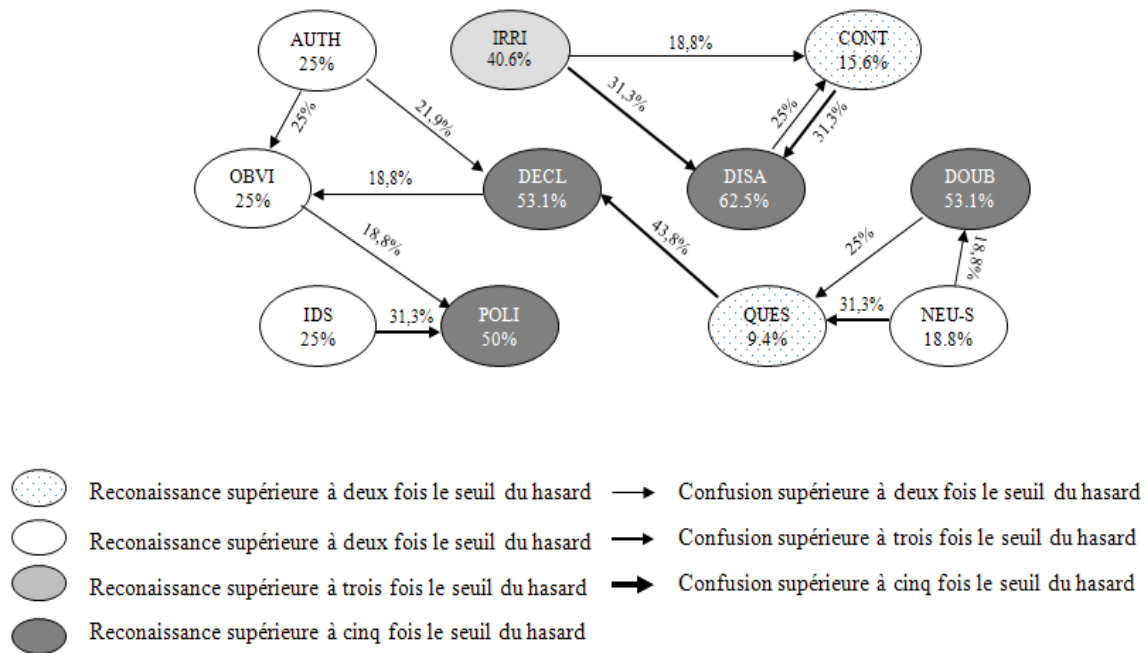


Figure 41: Confusions des 11 affects sociaux chinois pour 32 sujets natifs dans la modalité « vidéo seul ». Les pourcentages d'identification sont indiqués dans les ovales et les taux de confusions sur les flèches. Seules les confusions qui affichent des scores supérieurs à deux fois le seuil du hasard sont rapportées.

En ce qui concerne le regroupement des affects sociaux sur la base de leur proximité dans la modalité visuelle seule, la figure 42 illustre les résultats de la classification hiérarchique. D'après la graphique de perte d'inertie en fonction du nombre de groupes, il est raisonnable de considérer 3 catégories d'affects sociaux dans cette condition expérimentale.

La première catégorie regroupe la *déclaration* (DECL), la *question* (QUES), le *maternel* (IDS), la *politesse* (POLI), l'*autorité* (AUTH) et l'*évidence* (OBVI). Au sein de cette catégorie, les expressions de *déclaration* et de *question* sont plus particulièrement confondues sur la base des indices visuels. Un même constat est fait pour le *maternel* et la *politesse*, ainsi que pour l'*autorité* et l'*évidence*. La deuxième catégorie est composée de l'exclamation de *surprise neutre* (NEU-S) et du *doute* (DOUB). Bien que ces deux attitudes montrent des confusions importantes avec la *question*, cette dernière se rapproche plus de la *déclaration* (43,8%) (voir figure 41). Enfin, la troisième catégorie contient le *mépris* (CONT), la *déception* (DISA) et l'*irritation* (IRRI).

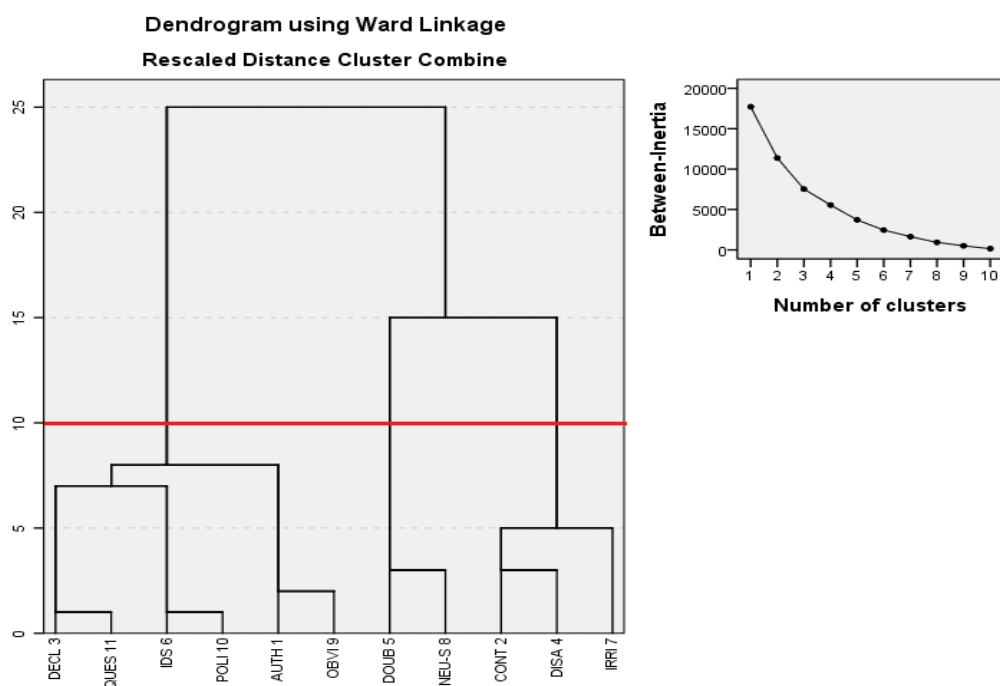


Figure 42: Classification hiérarchique des 11 affects sociaux chinois pour 32 sujets natifs dans la modalité « vidéo seul ».

- Modalité « audio-vidéo »

La figure 43 représente les taux d'identification et les plus importantes confusions des affects sociaux dans la modalité audiovisuelle. Par rapport aux modalités « audio seul » et « vidéo seul », la modalité audiovisuelle véhicule plus d'informations expressives du fait des synergies entre ces deux modalités. Par conséquent, les sujets reconnaissent mieux les affects sociaux dans cette condition expérimentale que dans les deux conditions séparées. On observe donc naturellement moins de confusions dans cette modalité. Presque tous les affects sociaux sont reconnus cinq fois au-dessus du hasard sauf le *mépris* (CONT), qui est cependant correctement reconnu (28,1%). La *déception* (DISA) reçoit le meilleur score de reconnaissance (90,6%).

Cependant, la *déclaration* (DECL) et l'*évidence* (OBVI) montrent des confusions réciproques, ainsi que le *doute* (DOUB) et la *question* (QUES). L'*autorité* (AUTH) est parfois confondue avec l'*évidence* (OBVI), de même que la *politesse* (POLI) avec la *déclaration* (DECL) et l'*irritation* (IRRI) avec le *mépris* (CONT). La *déception* (DISA), le *maternel* (IDS) et l'exclamation de *surprise neutre* (NEU-S) sont bien identifiés et sans confusions particulières.

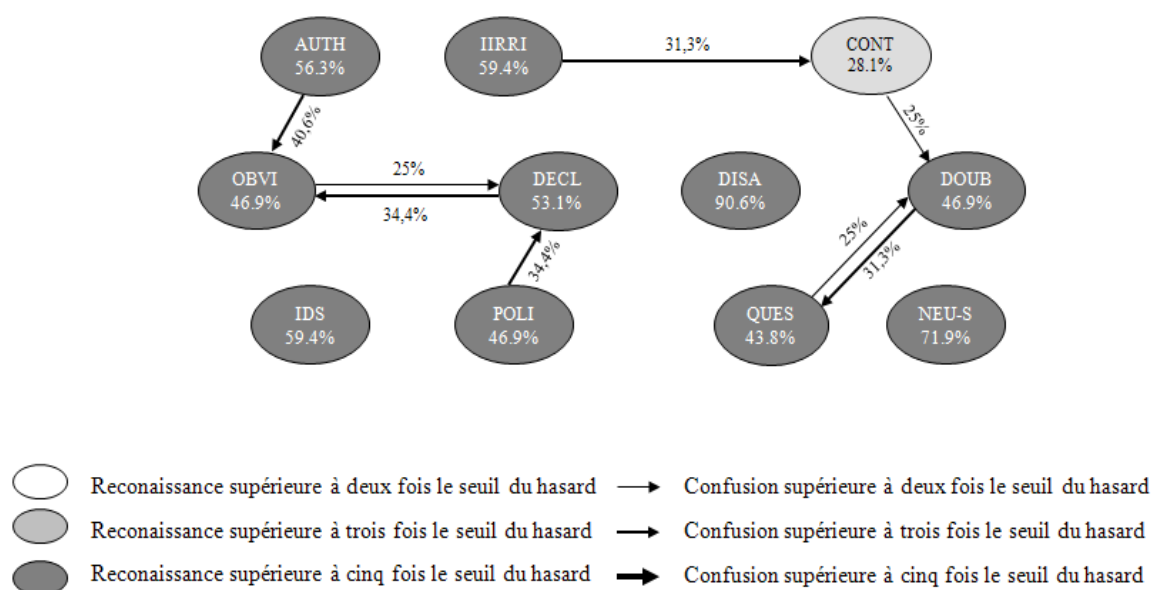


Figure 43: Confusions des 11 affects sociaux chinois pour 32 sujets natifs dans la modalité « audio-vidéo ». Les pourcentages d'identification sont indiqués dans les ovales et les taux de confusions sur les flèches. Seules les confusions qui affichent des scores supérieurs à deux fois le seuil du hasard sont rapportées.

Les résultats de la catégorisation de ces affects sociaux dans la condition « audio-vidéo » sont illustrés par la figure 44. D'après le graphique de perte d'inertie en fonction du nombre de groupes, les 11 affects sociaux se répartissent en 3 catégories pour cette condition expérimentale.

La première catégorie contient la *déclaration* (DECL), l'*évidence* (OBVI), la *politesse* (POLI), le *maternel* (IDS) et l'*autorité* (AUTH). Ces affects sociaux varient par leur force illocutoire (la *déclaration* et l'*évidence*) ou par leur contexte social (la *politesse*, le *maternel* et l'*autorité*). La deuxième catégorie regroupe le *doute* (DOUB), la *question* (QUES) et l'exclamation de *surprise neutre* (NEU-S) avec le *mépris* (CONT) et l'*irritation* (IIRRI). L'expression de *mépris* se rapproche de la *question* et du *doute*, ce qui a déjà été observé dans la modalité « audio seul ». La *déception* (DISA), bien reconnue, forme un isolat dans cette condition expérimentale.



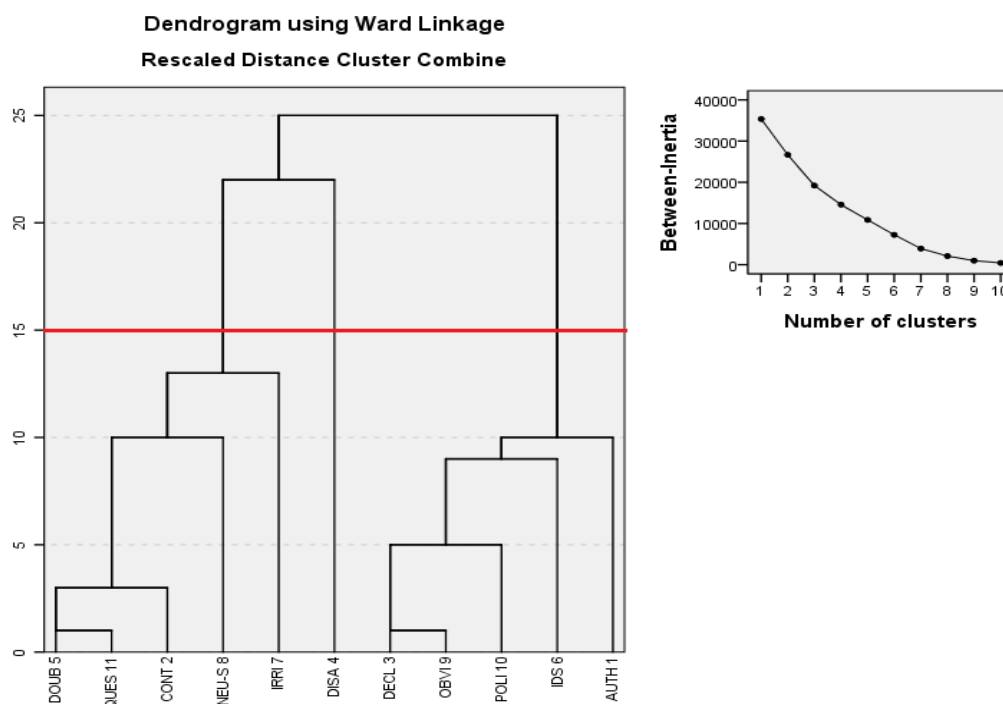


Figure 44: Classification hiérarchique des 11 affects sociaux chinois pour 32 sujets natifs dans la modalité « audio-vidéo ».

### 2.5.7. Conclusion

Cette étude de la perception multimodale des expressions socio-affectives est basée sur onze affects sociaux chinois sélectionnés parmi les 19 selon des critères de prototypicité. Un même énoncé exprimé avec les onze affects sociaux dans trois modalités (« audio seul », « vidéo seul » et « audio-vidéo »), ont été jugés par 32 sujets natifs.

Dans un premier temps, nous avons vérifié l'importance des facteurs sur la réponse des sujets grâce à une analyse de régression logistique mixte. Selon le résultat, les facteurs significatifs sont l'affect social, la modalité et l'ordre de présentation des modalités, ainsi que les interactions entre l'affect social et la modalité, et entre l'affect social et l'ordre de présentation des modalités. Le genre du sujet n'a pas d'effet significatif sur la perception de ces expressions.

Ensuite, pour chaque affect social, des analyses de comparaison multiple ont été appliquées pour mettre en évidence les différences significatives des proportions de bonne réponse entre chacune des trois modalités, ainsi qu'entre les deux ordres de présentation des modalités. Quelques différences significatives sont observées pour

*l'autorité*, la *déception*, le *maternel*, l'exclamation de *surprise neutre*, la *politesse* et la *question*.

Pour *l'autorité* et le *maternel*, le taux d'identification en « vidéo seul » est significativement inférieur à celui obtenu en « audio-vidéo » ( $p < 0,05$ ) ; pour la *question*, le stimulus présenté en « vidéo seul » est significativement mieux reconnu que celui en « audio seul » ; pour l'exclamation de *surprise neutre*, les taux d'identification dans les conditions audio seule et audiovisuelle sont significativement supérieurs à celui de la condition visuelle seule ( $p < 0,01$ ). Les indices auditifs contribuent ainsi de façon majeure à la reconnaissance de la *question* et de l'exclamation de *surprise neutre*. Pour la *politesse*, le taux d'identification en modalité audio seule est significativement inférieur à ceux obtenus en modalité audiovisuelle et visuelle seule ( $p < 0,05$ ), ce qui montre l'importance de l'information visuelle dans la reconnaissance de l'expression. Au cas de la *déception*, le taux d'identification en modalité audiovisuelle est significativement supérieur à ceux en audio seul et en vidéo seul, ce qui implique la coopération essentielle des deux modalités dans cette expression affective. Le cas de la *déclaration* est un peu particulier, car les taux de reconnaissance en trois modalités sont presque identiques.

Concernant l'effet de l'ordre de présentation des modalités, la majorité des affects sociaux ne montre pas de différences significatives entre les deux ordres de présentation des modalités, à l'exception du *maternel* et de l'exclamation de *surprise neutre* ( $p < 0,05$ ).

Au vu des matrices de confusion des affects sociaux dans chaque condition expérimentale, nous avons observé d'une part les principales confusions perceptives entre les affects sociaux ainsi que l'orientation de ces confusions, et d'autre part les tendances générales du regroupement de ces affects sociaux selon leur proximité. L'expression de *politesse* est largement confondue avec la *déclaration* en condition audio seule (53,6%) ; la *question* n'est pas bien reconnue dans la condition visuelle seule, et est principalement confondue avec la *déclaration* (43,8%).

Selon les résultats d'une analyse de classification hiérarchique sur les mêmes matrices de confusion, et pour chaque modalité de l'expression attitudinale, les onze affects sociaux sont regroupés dans des catégories en fonction des modalités considérées.

## 2.6. Conclusion du chapitre

En nous référant à différents travaux menés dans le cadre de recherches sur les attitudes dans différentes langues (Grépillat, 1996; Aubergé et al., 1997; Morlec, 1997; Diaferia, 2002; Rilliard et al., 2008; Shochi, 2008; de Moraes et al., 2010; Mac et al., 2010), nous avons sélectionné 19 affects sociaux réalisés en chinois mandarin, qui sont usuels dans la vie courante des chinois, pour des études de la perception de la prosodie des affects sociaux, à long terme, son apprentissage auprès d'apprenants étrangers.

Un corpus composé de 2888 stimuli, à savoir 19 affects sociaux exprimés sur 152 énoncés, a été enregistré et évalué sur une série d'expériences perceptives. Différentes sources de variation prosodique ont été prises en compte dans la conception de ce corpus de parole: la longueur des phrases (en syllabes), le placement des tons et la structure syntaxique des énoncés sont ainsi systématiquement variés.

Par la suite, le corpus a été évalué par des auditeurs chinois lors d'une expérience perceptive pendant laquelle 399 expressions socio-affectives sélectionnées doivent être reconnues en audio seul par les auditeurs.

Les résultats de l'analyse indiquent d'une part une reconnaissance globale supérieure au niveau du hasard de presque tous les affects sociaux sauf l'attitude de *confiance*. D'après le résultat de l'analyse de la variance, la longueur des énoncés et l'interaction entre longueur et affect social ont un effet significatif ( $p < 0,01$ ) sur la reconnaissance des affects sociaux. Les énoncés monosyllabiques reçoivent les taux d'identification les plus bas et se différencient significativement des autres longueurs ( $p < 0,01$ ). Le genre a aussi un effet significatif ( $p < 0,01$ ) mais de peu d'ampleur par rapport aux autres facteurs. Les femmes (27,2%) reconnaissent généralement mieux les affects sociaux que les hommes (22,8%). En matière de confusions perceptives entre les affects sociaux, la *déclaration* attire la plupart des confusions alors qu'elle est elle-même bien identifiée (58,4%) ; la *déception* est aussi bien reconnue mais peu confondue avec les autres affects sociaux. Une analyse de classification hiérarchique regroupe les affects sociaux en sept catégories. Les expressions d'exclamation de *surprise*, de *doute* et de *question* ont sont séparées de l'ensemble des autres affects sociaux. Cela montre la cohérence entre perception et expression de ces attitudes, au travers desquelles la locutrice interroge son

interlocuteur, notamment à l'aide d'une intonation montante en fin de phrase. La même constatation a été faite par Shochi (2008) sur les attitudes japonaises.

Une analyse de classification hiérarchique a été menée sur les composantes principales d'une ACP, elle-même menée sur 10 indices statistiques des trois paramètres prosodiques (F0, intensité, durée), extraits des affects sociaux. Les résultats nous permettent d'identifier les traits prosodiques principaux des affects sociaux :

- *L'admiration*, le *maternel*, le *mépris* et le *doute* sont caractérisés par une durée moyenne plus longue de la première et de la dernière syllabe et donc de l'ensemble de l'énoncé, de même que par des valeurs élevées de tous les indices de F0 ;
- *L'autorité*, la *déclaration*, la *résignation* et la *déception* sont au contraire caractérisées par une durée moyenne courte notamment pour la dernière syllabe, et également par un registre bas, une plage étroite de F0 et moins de variabilité de F0 ;
- *L'évidence*, la *séduction*, la *politesse* et la *confiance* sont réalisées avec des caractéristiques de F0 proches de celles du groupe précédent, mais elles présentent une durée plus importante.
- *L'ironie* et *l'intimité* se caractérisent par une durée moyenne très courte et donc une énonciation rapide, ainsi qu'une importante pente descendante de F0 sur l'ensemble de l'énoncé;
- La *surprise positive*, la *surprise négative*, la *surprise neutre*, la *question* et *l'irritation* sont réalisées avec une large plage de F0, un niveau élevé de F0 moyenne pour la première et la dernière syllabe ainsi que pour l'ensemble de l'énoncé, et également par une durée moyenne élevée de la dernière syllabe.

Par la suite, une analyse des contours intonatifs du même corpus a été effectuée. Les contours intonatifs des affects sociaux ont été comparés l'un avec l'autre en terme de RMS de F0 et de corrélation de la forme. Les résultats montrent que les 19 affects sociaux peuvent être répartis en différentes catégories, selon la longueur des énoncés, et sur la base de ces mesures. En première analyse, deux catégories principales sont observées comme dans le test perceptif de ces mêmes affects sociaux chinois auprès de sujets natifs. La plupart des regroupements d'affects sociaux sont similaires, en dépit des

variations de longueur des énoncés, sur la base de ces deux mesures de distance objective.

Les résultats de l'analyse acoustique nous permettent également de comprendre certaines des confusions perceptives entre affects sociaux, comme la confusion entre la *déclaration* et l'*évidence* et la *confiance*, celle entre l'*irritation* et l'*autorité*, et celle entre la *déception* et la *résignation*. Cependant, certaines confusions observées peuvent avoir des causes plus complexe et ne peuvent pas s'expliquer uniquement par des similarités d'indices acoustiques : c'est le cas de la confusion entre le *maternel* et la *séduction*.

Dans le deuxième test perceptif, onze affects sociaux ont été sélectionnés afin de maximiser la prototypicité, à des fins d'usage didactique de la prosodie des affects sociaux en français et en chinois langue étrangère ou seconde. Un seul énoncé a été sélectionné pour le test perceptif. Il a été exprimé avec ces onze affects sociaux et a été présenté dans trois modalités (soit 33 stimuli). Les principaux objectifs de ce test sont d'évaluer la reconnaissance des affects sociaux chinois présentés dans différentes modalités auprès de sujets natifs, et d'examiner l'effet de la modalité, de l'ordre de présentation des modalités, ainsi que du genre des sujets sur la reconnaissance de ces affects sociaux.

Tout d'abord, une analyse de régression logistique mixte a été effectuée pour évaluer la significativité d'un ensemble de facteurs sur les réponses des sujets. Le résultat montre que la modalité des expressions socio-affectives et l'ordre de présentation des modalités ont des effets significatifs sur la reconnaissance de certains affects sociaux. Une analyse de comparaisons multiples permet de mieux comprendre ces effets: l'*autorité*, le *maternel* et la *question*, dont le taux de reconnaissance en modalité audiovisuelle est significativement supérieur à celui obtenu en modalité visuelle seule (l'*autorité* et le *maternel*) et en modalité audio seule (la *question*); l'exclamation de *surprise neutre*, dont les taux d'identification dans les conditions audio seule et audiovisuelle sont significativement supérieurs à celui observé pour la condition visuelle seule ; la *politesse* pour laquelle les taux de reconnaissance en modalité visuelle seule et audiovisuelle sont significativement plus élevés qu'en modalité audio seule.

Il est également intéressant de remarquer que les différentes modalités n'ont pas le même rôle dans la reconnaissance des expressions socio-affectives (pour des résultats

concordants, voir de Moraes & Rilliard, 2014). Les expressions d'exclamation de *surprise neutre* et de *question* sont plus dépendantes de l'information auditive ; celle de *politesse* est mieux reconnue lorsque les informations visuelles sont présentes; pour les expressions de *mépris*, de *doute*, de *déception*, d'*évidence* et d'*irritation*, les sujets ont plutôt recours aux informations auditives et visuelles combinées. En général, la synergie des modalités auditive et visuelle permet aux sujets de mieux reconnaître les affects sociaux qu'une seule modalité dans la plupart des cas. Cependant, il semble que les différentes modalités n'exercent pas d'influence particulière sur la reconnaissance de la *déclaration*, qui reçoit de bons scores de reconnaissance dans les trois conditions expérimentales.

Ensuite, une analyse des confusions entre les affects sociaux permet d'observer les similarités perceptives entre ces onze affects sociaux dans chacune des trois modalités. Notons que des confusions de la *politesse* vers la *déclaration* et de l'*autorité* vers l'*évidence* sont observées dans les trois modalités, et que la *question* exprimée en condition visuelle seule n'est pas reconnue par les sujets, et est confondue principalement avec la *déclaration*. Cela pourrait être expliqué par le fait que la *déclaration* et la *question*, deux modalités de base de l'énonciation, sont considérées comme affectivement neutres. Par conséquent, lorsque la locutrice exprime une déclaration et une question, elle ne manifeste pas de gestualité corporelle ni d'expression faciale spécifique, donc en l'absence d'information prosodique, les sujets ne parviennent pas à juger l'expression. En considérant l'ensemble de confusions observées lors de l'expérience, nous avons pu regrouper les stimuli perçus dans chaque modalité selon la proximité des distributions des confusions pour chaque affect social proposé. Les onze affects sociaux sont regroupés dans des catégories différentes en fonction des modalités considérées. Il est à noter que la *question* est classée avec la *déclaration* par les sujets natifs en modalité visuelle seule, mais avec le *doute* dans les modalités audio seule et audiovisuelle.

# **Chapitre 3. Etude interculturelle de la prosodie socio-affective en chinois et des concepts des affects sociaux**

## **3.1. Introduction**

Dans le chapitre précédent, nous avons évalué la pertinence du corpus prosodique des affects sociaux en chinois mandarin et nous avons examiné la perception de ces affects sociaux à partir de la prosodie chez des auditeurs natifs en présentant les variations d'identification et les confusions. Dans ce présent chapitre, nous nous pencherons sur la perception auditive et le traitement cognitif de ces affects sociaux à partir des études interculturelles.

Nous avons suivi l'hypothèse que l'expression des affects sociaux est produite dans un contrôle volontaire et est acquise pendant l'enfance dans l'environnement socioculturel natif. Ainsi, les concepts des socio-affects eux-mêmes, et les expressions prosodiques de ces affects sociaux peuvent varier d'une culture/langue à l'autre. De mauvaises interprétations de l'intention du message pourraient être donc causées par ces différences culturelles (Pavlenko, 2007). Dans ce cas, une étude contrastive interculturelle permettra de mettre en évidence les similarités potentiellement universelles et les spécificités perceptives des affects sociaux chez des auditeurs de langue maternelle différente (Kramer, 1964 ; Van Bezooijen et al., 1983 ; Fónagy, 1983 ; Mejvaldová, 2000 ; Shochi et al., 2007).

Par ailleurs, en tant que langue tonale, le chinois mandarin est caractérisé par quatre tons lexicaux, plus un ton neutre. Ces tons constituent la difficulté majeure pour des apprenants de langue maternelle non tonale comme le français. Ainsi Mac (2012) a montré que les tons vietnamiens perturbent la perception des affects sociaux par des français, même s'il semble qu'ils sont capables de reconnaître la perturbation locale des tons en regard de la perturbation globale de la prosodie des affects sociaux. Nous supposons donc que les tons chinois perturberont la perception des auditeurs français lors d'une tâche d'identification des affects sociaux proposés avec les mêmes

descripteurs pour le chinois. Pour vérifier cette hypothèse, nous prenons un troisième groupe de référence, précisément des auditeurs vietnamiens, car par rapport à la langue française, la langue vietnamienne est linguistiquement et culturellement plus proche de la langue chinoise avec des similarités observées entre elle et le chinois (Michaud, 2008). Par conséquent, supposant un effet du ton sur la reconnaissance des affects sociaux chinois, comparer le comportement perceptif des auditeurs français avec celui des auditeurs vietnamiens comporte moins de biais que de comparer directement les auditeurs français avec les auditeurs natifs.

Par ailleurs, pour chaque langue, la nature même des concepts des affects sociaux ainsi que l'interprétation faite par les sujets de l'étiquette de chaque affect et leur interprétation de la description contextuelle donnée constituent également un aspect non négligeable et incontournable de cette étude. Au XX<sup>ème</sup> siècle dans les années 50, Sapir et Whorf ont supposé la dépendance de la vision du monde de l'individu aux catégories linguistiques, autrement dit, la façon dont on voit le monde dépendrait de notre langue. C'est le principe du relativisme linguistique : la structure d'une langue influence la manière dont les tenants conceptualisent le monde, ainsi que leur processus cognitif. Or, Berlin & Kay (1969) ont remis en question cette hypothèse. Ils ont proclamé que la terminologie des couleurs est sujette à des contraintes sémantiques universelles et que la cognition ou la perception de chaque catégorie de couleur est aussi universelle. Leur point de vue a provoqué un grand débat sur l'universalisme de la terminologie des couleurs entre les défenseurs d'une vision universaliste (par exemple : Bornstein et al., 1976, Kay & Maffi, 1999) et ceux d'une vision relativiste (par exemple: Saunders, 1995, Lucy, 1997, Wierzbicka, 2006). La traduction de termes d'une langue à une autre a été vivement critiquée chez Berlin & Kay (1969) par les relativistes. Pour éviter les contraintes sémantiques et culturelles, Wierzbicka a proposé d'éliminer les termes spécifiques à une langue en utilisant une métalangue sémantique naturelle (*Natural Semantic Metalanguage en anglais*) pour représenter un concept (Wierzbicka, 1992, 1996, 1999, 2005).

Beaucoup d'autres études menées dans divers domaines et inspirées par celles de Berlin et Kay, se sont attachées à examiner la relation entre la représentation sémantique et la conceptualisation du monde. Parmi elles, les travaux de Brazill et al. (1995), Romney et al. (1995, 1996, 1997), Romney & Moore (1998), Moore et al. (1999), ont été réalisés au



sujet de noms d'animaux, de termes émotionnels et de termes de parenté. En ce qui concerne les « concepts » d'affects sociaux, ils ont aussi besoin d'être mis en regard des études perceptives qui mesurent les possibles (vraies ou fausses) ressemblances de la morphologie acoustico-perceptive et les différences de représentation conceptuelle.

Dans notre étude sur l'enseignement de la prosodie des affects sociaux en langues étrangères ou secondes, nous rencontrons le même problème concernant les concepts d'affects sociaux : sont-ils universels (dans notre étude, nous nous intéressons uniquement aux langues chinoise et française) ? Dans une langue et culture, la catégorisation conceptuelle des affects sociaux converge-t-elle vers la catégorisation perceptive ? De même, afin d'atténuer les potentielles contraintes sémantiques, et au lieu d'utiliser une métalangue neutre comme Wierzbicka, nous recourrons à une description de la situation communicative pour tous les affects sociaux étudiés. Cette façon de représenter les affects repose sur le postulat selon lequel les affects sociaux sont souvent liés aux relations dynamiques entre les individus et aux situations où ils s'expriment (Seighworth & Gregg, 2010; Seyfert, 2012). Par ailleurs, les étiquettes d'affects sociaux utilisées pour représenter leurs concepts sont des étiquettes vernaculaires dans une langue donnée.

## **3.2. Perception interculturelle des affects sociaux chinois**

### **3.2.1. Introduction**

Après avoir validé le corpus des affects sociaux chinois avec des sujets natifs, nous étudions au sein de cette section le comportement perceptif des sujets français et des sujets vietnamiens pour les mêmes expressions socio-affectives en chinois mandarin. Cette étude s'attache à examiner la reconnaissance des affects sociaux par des sujets français et vietnamiens et l'effet potentiel du ton sur la reconnaissance. Nous observons aussi les confusions perceptives et le classement des affects sociaux par les sujets de chaque langue.

### **3.2.2. Sous-corpus**

Nous avons sélectionné pour le test perceptif 304 stimuli (16 énoncés \* 19 affects sociaux) sur un seul critère de valeur et de position de tons lexicaux. Il n'y a aucune

variation morpho-syntaxique et sémantique dans le sous-corpus, puisque tous les énoncés sont des groupes nominaux ou verbaux, ou des chiffres.

Cette sélection contient quatre mots de trois syllabes et douze énoncés de quatre syllabes. Pour les énoncés à quatre syllabes et un énoncé à trois syllabes, les quatre tons sont systématiquement variés sur la première et la dernière syllabe, les autres syllabes sont porteuses d'un ton plat. Pour les trois autres mots trisyllabiques, le ton neutre est utilisé sur la dernière syllabe. La table 14 présente les énoncés sélectionnés pour le sous-corpus. Les chiffres dans la colonne Tons représentent la combinaison des tons de l'énoncé : 1 fait référence à un ton plat, 2 à un ton montant, 3 à un ton descendant-remontant, 4 à un ton descendant et 0 à un ton neutre.

### 3.2.3. Sujets

Etant donné qu'un des objectifs de cette expérience perceptive porte sur l'influence potentielle des tons sur la perception des auditeurs français naïfs, pour avoir une vision plus objective, nous avons sélectionné comme groupe de référence des auditeurs vietnamiens naïfs en chinois qui sont culturellement et linguistiquement plus proches des chinois.

Ainsi, 15 auditeurs français (9 hommes et 6 femmes) d'âge moyen de 33 ans, et 15 auditeurs vietnamiens (7 hommes et 8 femmes) d'âge moyen de 27 ans, ont participé à cette expérience. Ils habitent tous dans l'agglomération grenobloise et aucun n'a signalé de trouble auditif.

### 3.2.1. Protocole expérimental

Le test perceptif s'est déroulé via une interface informatique, réalisée avec le logiciel « LiveCode ». Lors de la passation, le sujet était assis confortablement face à l'écran d'ordinateur dans un lieu calme, et tous les stimuli lui étaient présentés à travers un casque audio de haute qualité. Le test débutait par une présentation globale et une description de chaque affect social illustré par des exemples de contextes possibles. Ces informations étaient écrites dans la langue maternelle du sujet (en français pour les sujets français et en vietnamien pour les sujets vietnamiens). Les sujets conservaient ce document comme référence pendant le déroulement de test. Après avoir lu la consigne du test, les sujets commençaient le test. Chaque stimulus était présenté une seule fois, dans un ordre aléatoire. Pour chacun, les sujets doivent juger quel est l'affect social

exprimé par la locutrice en choisissant une étiquette parmi les 19 proposées à l'écran (les étiquettes sont écrites en français). Aucune limitation de temps de réflexion avant la présentation du stimulus suivant n'est imposée. Les réponses données ont été sauvegardées automatiquement sur le disque dur de l'ordinateur.

Nb de syllable	Ton	Phrase en chinois	Transcription en pinyin	Traduction française
3	2-2-0	胡萝卜	hu2luo2bo	carotte
	2-3-0	量尺寸	liang2chi3cun	mesurer la taille
	4-4-0	做记号	zuo4ji4hao	faire une marque
	1-1-1	出租车	chu1zu1che1	taxi
4	1-1-1-2	交通规则	jiao1tong1gui1ze2	Le code routier
	2-1-1-2	长沙中学	chang2sha1zhong1xue2	le lycée de Changsha
	3-1-1-2	北京中学	bei3jing1zhong1xue2	le lycée de Beijing
	4-1-1-2	四双新鞋	si4shuang1xin1xie2	quatre nouvelles paires de chaussures
	1-1-1-3	三支铅笔	san1zhi1qian1bi3	trois crayons
	2-1-1-3	十只鹦鹉	shi2zhi1ying1wu3	dix perroquets
	3-1-1-3	五千一百	wu3qian1yi1bai3	5100
	4-1-1-3	四千三百	si4qian1san1bai3	4300
	1-1-1-4	公交车站	gong1jiao1che1zhan4	l'arrêt de l'autobus
	2-1-1-4	环山公路	huan2shan1gong1lu4	la route de montagne
	3-1-1-4	五千只鹤	wu3qian1zhi1he4	5000 grues
	4-1-1-4	四天三夜	si4tian1san1ye4	quatre jours et trois nuits

Table 14: Sous-corpus utilisé dans l'expérience perceptive interculturelle de 19 affects sociaux chinois auprès de sujets français et vietnamiens.

### 3.2.2. Analyse des résultats

#### 3.2.2.1. Facteurs influant la reconnaissance des affects sociaux

Dans un premier temps, une analyse de variance (ANOVA) à trois facteurs aléatoires a été réalisée sur les réponses des sujets. Les trois facteurs fixés sont l'affect social présenté (Affect social, 19 niveaux), la combinaison des tons (Ton, 16 niveaux) et la langue maternelle des sujets (Langue, 2 niveaux). Cette analyse vise à évaluer l'impact de ces facteurs sur la qualité des scores de reconnaissance des sujets. Le niveau de significativité ( $p$ ) est fixé à 0,05.

D'après les résultats de l'ANOVA rapportés dans la table 15, les facteurs d'affect social et de ton ont un effet significatif sur la reconnaissance des affects sociaux ( $p < 0,01$ ). De plus, leur interaction ainsi que celle entre l'affect social et la langue maternelle des sujets sont aussi significatifs ( $p < 0,01$ ). L'effet significatif de l'affect social explique la plus grande part de la variance observée ( $\eta^2 = 0,070$ ), avec l'interaction entre les facteurs d'affect social et de ton ( $\eta^2 = 0,060$ ), ce qui montre l'importance de leur influence sur la réponse de sujet. Le facteur de ton ainsi que l'interaction entre l'affect social et la langue maternelle des sujets sont significatifs ( $p < 0,01$ ), mais n'expliquent qu'une part mineure de la variance. La langue maternelle des sujets, l'interaction entre les facteurs de langue et de ton, ainsi que celle entre les trois facteurs, n'ont cependant pas montré d'effet significatif ( $p > 0,05$ ).

	SCE	ddl	F	<i>p</i>	$\eta^2$
<b>Affect social</b>	<b>81,85</b>	<b>18</b>	<b>35,8</b>	<b>0,000</b>	<b>0,070</b>
Langue	0,11	1	0,8	0,362	0,000
<b>Ton</b>	<b>4,61</b>	<b>15</b>	<b>2,4</b>	<b>0,001</b>	<b>0,004</b>
<b>Affect social: Langue</b>	<b>10,36</b>	<b>18</b>	<b>4,5</b>	<b>0,000</b>	<b>0,009</b>
<b>Affect social: Ton</b>	<b>68,53</b>	<b>270</b>	<b>2,0</b>	<b>0,000</b>	<b>0,060</b>
Langue : Ton	3,11	15	1,6	0,058	0,003
Affect social: Langue : Ton	37,72	270	1,1	0,131	0,034

Table 15: Effets des facteurs *Affect social*, *Langue* et *Ton* sur la reconnaissance des affects sociaux chinois par des auditeurs français et vietnamiens par un test ANOVA (effets significatifs en gras).

Afin d'estimer la significativité des facteurs d'affect social et de ton pour la reconnaissance des affects sociaux chinois propre à chaque groupe d'auditeurs (français et vietnamiens), nous avons effectué deux ANOVAs séparées. Le résultat présenté dans la table 16 montre que le facteur d'affect social a un effet significatif ( $p < 0,01$ ) ; le facteur de ton est significatif pour la perception des sujets français ( $p < 0,01$ ), alors qu'il ne l'est pas chez les sujets vietnamiens ( $p > 0,05$ ). L'interaction entre les facteurs d'affect social et de ton montre un effet significatif dans la perception des deux groupes de sujets ( $p < 0,01$ ).

	ddl	Français		Vietnamiens	
		F	p	F	p
<b>Affect social</b>	<b>18</b>	<b>20,2</b>	<b>0,000</b>	<b>20,1</b>	<b>0,000</b>
<b>Ton</b>	15	2,4	0,002	1,6	0,054
<b>Affect social : Ton</b>	<b>270</b>	<b>1,4</b>	<b>0,000</b>	<b>1,4</b>	<b>0,000</b>

Table 16: Résultat des ANOVAs séparées par langue maternelle des sujets pour la reconnaissance des affects sociaux chinois (effets significatifs en gras).

### 3.2.2.2. Taux d'identification des affects sociaux

La figure 45 représente les taux de reconnaissance de chaque affect social. Nous pouvons y constater que presque tous les affects sociaux sont reconnus au dessus du hasard, à l'exception de l'*ironie* (IRON) (2,7%). La *déception* (DISA) (37,3%) est l'attitude la mieux reconnue parmi les 19.

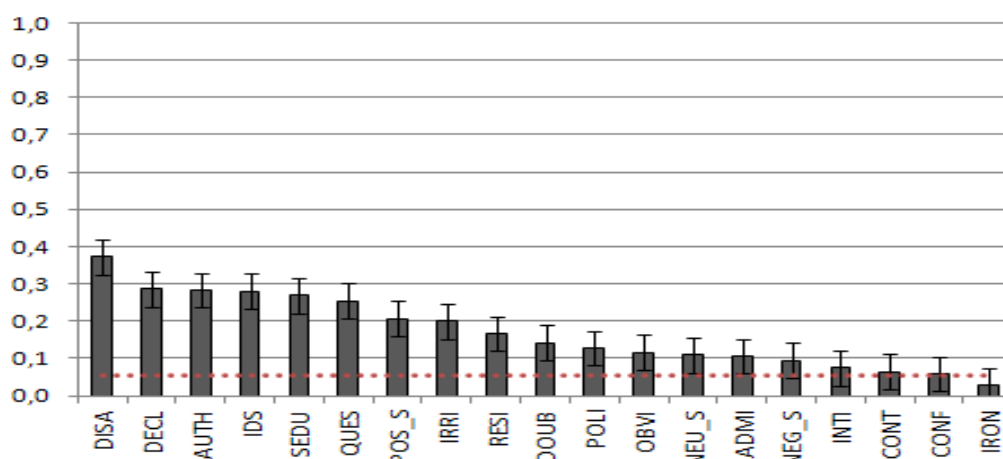


Figure 45: Taux d'identification des 19 affects sociaux chinois pour 30 sujets français et vietnamiens. Les affects sociaux sont présentés en abscisse et les scores de reconnaissance en ordonnée. La ligne rouge pointillée indique le seuil du hasard (5,3%).

### 3.2.2.3. Effet de la langue maternelle des sujets

La figure 46 représente les taux d'identification des 19 affects sociaux en fonction de la langue maternelle des sujets. Presque tous les affects sociaux sont reconnus au dessus du hasard, à l'exception du *mépris* (CONT) (3,8%), de la *confiance* (CONF) (3,3%) et de l'*ironie* (2,1%) pour les auditeurs français. Les auditeurs vietnamiens ont aussi un taux de reconnaissance de l'*ironie* (3,3%) inférieur au hasard.

En général, les auditeurs français reconnaissent mieux la *déclaration* (DECL) (33,8%), alors que les auditeurs vietnamiens perçoivent mieux la *déception* (DISA) (46,7%). Les sujets français identifient mieux la *séduction* (SEDU), l'*autorité* (AUTH), l'*évidence* (OBVI), la *déclaration* (DECL), la *question* (QUES) et le *doute* (DOUB) que les sujets vietnamiens, qui reconnaissent contrairement mieux la *déception* (DISA), l'*irritation* (IRRI), la *politesse* (POLI), le *mépris* (CONT) et la *confiance* (CONF). Cependant, l'*admiration* (ADMI), le *maternel* (IDS) et l'*intimité* (INTI) montrent des scores de reconnaissance similaires pour les deux langues maternelles, alors que l'*ironie* (IRON) est très mal identifiée par les deux groupes. Ce dernier point rappelle l'observation de Shochi (2008) et de Mac (2012), qui remarquaient également que l'*ironie* exprimée en français (Shochi, 2008) ou en vietnamien (Mac 2012), était l'attitude la plus difficile à identifier par les auditeurs français. Leurs observations et nos premiers résultats suggèrent que la reconnaissance de l'*ironie* pourrait dépendre fortement du contexte de communication. Cette hypothèse demanderait à être approfondie et testée par d'autres études.

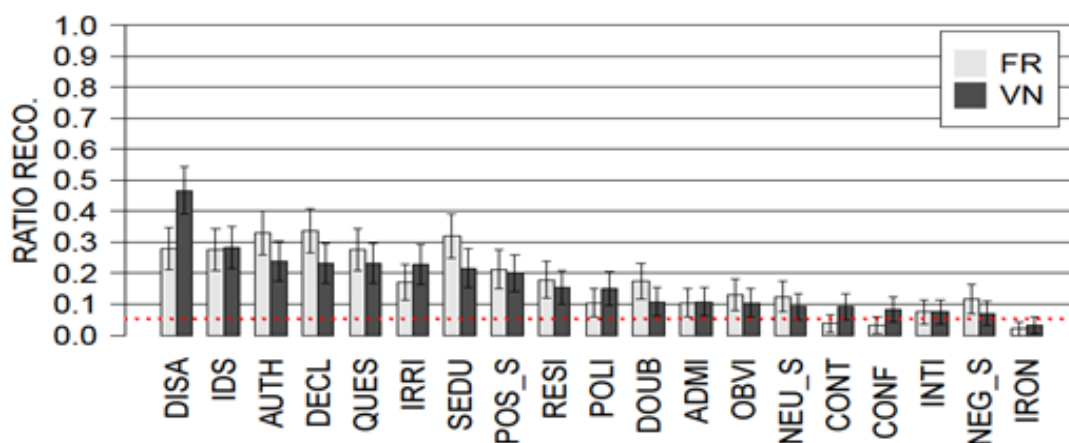


Figure 46: Taux d'identification des 19 affects sociaux chinois par 15 sujets français (FR) et 15 sujets vietnamiens (VN). Les affects sociaux sont présentés en abscisse et les scores de reconnaissance en ordonnée. La ligne rouge pointillée indique le seuil du hasard (5,3%).

#### 3.2.2.4. Effet du ton

D'après Le résultat des analyses de variance détaillées ci-dessus (voir partie 3.2.5.1), l'effet du facteur de combinaison de tons sur la reconnaissance des affects sociaux n'est significatif que chez les sujets français, son interaction avec le facteur d'affect social a un effet significatif pour tous les sujets, quelle que soit leur langue maternelle. La figure 47 représente les taux de reconnaissance des 19 affects sociaux pour chaque énoncé.

Globalement, la majorité des scores de reconnaissance est entre 15% et 20% pour tous les auditeurs. Pour les sujets français, les énoncés dont la structure de tons est 4-1-1-3 et 4-1-1-4 ont été les mieux reconnus avec des scores supérieurs à 20% (23,9% et 21,1%). Pour les auditeurs vietnamiens, la combinaison des tons 1-1-1-2 est celle la plus souvent identifiée correctement (21,1%). L'énoncé 2-2-0 recueille le score le plus bas tant pour les sujets français (9,5%) que pour les sujets vietnamiens (11,5%).

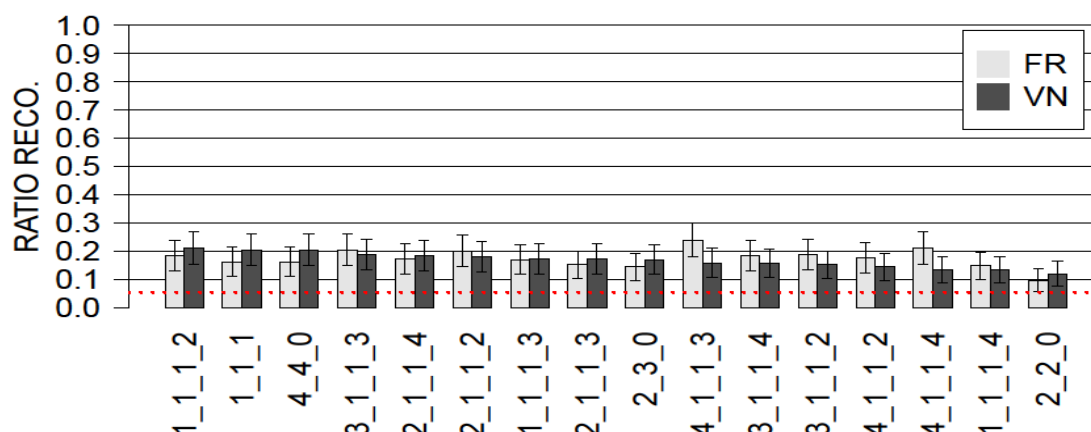


Figure 47: Taux d'identification des 19 affects sociaux chinois pour chaque énoncé par 15 sujets français (FR) et 15 sujets vietnamiens (VN). Les énoncés, présentés en abscisse, sont représentés par leur séquence tonale (1 représente le ton plat, 2 le ton montant, 3 le ton descendant-remontant, 4 le ton descendant et 0 le ton neutre); les scores de reconnaissance sont présentés en ordonnée. La ligne rouge pointillée indique le seuil du hasard (5,3%).

Parallèlement, en observant les taux d'identification pour chaque combinaison des tons en fonction des 19 affects sociaux et de chaque groupe de sujets, nous remarquons que la reconnaissance de certains affects sociaux varie en fonction du ton final de l'énoncé. La table 17 présente les taux de reconnaissance des 19 affects sociaux pour les énoncés dont la dernière syllabe porte successivement les quatre tons et le ton neutre. Le taux moyen de reconnaissance des affects sociaux chinois pour ces combinaisons des tons y est aussi rapporté au bas de chaque colonne.

	2_2_0		2_3_0		4_4_0		1_1_1		1_1_1_2		1_1_1_3		1_1_1_4	
	Fr	Vn	Fr	Vn	Fr	Vn	Fr	Vn	Fr	Vn	Fr	Vn	Fr	Vn
ADMI	6,7	0,0	13,3	6,7	33,3	40,0	0,0	0,0	0,0	6,7	26,7	6,7	6,7	0,0
AUTH	13,3	20,0	40,0	33,3	60,0	6,7	26,7	40,0	60,0	40,0	6,7	13,3	26,7	20,0
CONF	0,0	0,0	13,3	0,0	0,0	20,0	0,0	6,7	0,0	6,7	0,0	13,3	6,7	20,0
CONT	0,0	20,0	6,7	0,0	13,3	20,0	6,7	6,7	0,0	20,0	6,7	20,0	0,0	13,3
DECL	20,0	0,0	13,3	26,7	20,0	13,3	26,7	13,3	46,7	20,0	46,7	26,7	26,7	13,3
DISA	33,3	73,3	13,3	53,3	13,3	33,3	46,7	33,3	20,0	60,0	40,0	46,7	26,7	20,0
DOUB	0,0	13,3	20,0	6,7	20,0	0,0	6,7	13,3	20,0	26,7	6,7	6,7	13,3	0,0
IDS	0,0	0,0	20,0	40,0	20,0	20,0	46,7	53,3	26,7	66,7	20,0	33,3	26,7	20,0
INTI	6,7	6,7	6,7	6,7	6,7	13,3	6,7	6,7	0,0	20,0	6,7	6,7	26,7	6,7
IRON	0,0	0,0	0,0	13,3	0,0	0,0	0,0	0,0	6,7	0,0	6,7	0,0	6,7	0,0
IRRI	13,3	40,0	6,7	20,0	33,3	33,3	6,7	46,7	6,7	13,3	13,3	6,7	13,3	26,7
NEG-S	0,0	13,3	0,0	6,7	26,7	6,7	46,7	26,7	6,7	0,0	6,7	0,0	13,3	13,3
NEU-S	13,3	6,7	13,3	13,3	0,0	6,7	6,7	13,3	13,3	20,0	13,3	13,3	6,7	26,7
OBVI	0,0	6,7	13,3	0,0	13,3	13,3	0,0	0,0	20,0	13,3	13,3	26,7	13,3	13,3
POLI	13,3	0,0	6,7	13,3	6,7	26,7	0,0	6,7	13,3	0,0	6,7	40,0	0,0	6,7
POS-S	6,7	6,7	6,7	20,0	0,0	40,0	6,7	13,3	33,3	26,7	33,3	6,7	26,7	26,7
QUES	20,0	6,7	6,7	6,7	6,7	6,7	40,0	53,3	53,3	33,3	13,3	20,0	6,7	6,7
RESI	26,7	13,3	33,3	26,7	13,3	20,0	20,0	13,3	6,7	20,0	6,7	6,7	13,3	20,0
SEDU	6,7	0,0	40,0	26,7	20,0	26,7	13,3	40,0	13,3	6,7	46,7	33,3	20,0	0,0
moyen	9,5	11,9	14,4	16,8	16,1	18,2	16,2	20,4	18,2	21,1	16,9	17,2	14,7	13,3

Table 17: Taux d'identification des 19 affects sociaux pour les énoncés dont la dernière syllabe porte le ton 1, ton 2, ton 3, ton 4 ou le ton neutre, pour tous les sujets français et vietnamiens. Le taux moyen d'identification pour chaque énoncé est rapporté sur la dernière ligne.

Pour les énoncés constitués de ton haut et plat (ton1) sur les premières syllabes et où seul le ton final varie en fonction des quatre tons ou du ton neutre, les sujets français ont reconnu mieux l'exclamation de *surprise négative* (NEG-S) quand la dernière syllabe de l'énoncé porte un ton 1, mais mal reconnu l'exclamation de *surprise positive* (POS-S) produite par le même énoncé. La *question* (QUES) et l'*autorité* (AUTH) reçoivent le meilleur score quand le ton montant (ton 2) est positionné sur la dernière syllabe. Pour l'*admiration* (ADMI) et la *séduction* (SEDU), l'énoncé avec un ton descendant-remontant (ton 3) à la fin obtient un score de reconnaissance plus élevé que les autres énoncés, et l'*autorité* enregistre son score le plus faible. Les sujets français reconnaissent bien l'*intimité* (INTI) si le ton descendant (ton 4) se positionne à la fin de phrase, mais ils identifient mal la *question* produite par le même énoncé.

Les sujets vietnamiens reconnaissent bien l'*irritation* (IRRI), l'exclamation de *surprise négative* (NEG-S), la *question* (QUES) et la *séduction* (SEDU) quand la dernière syllabe de l'énoncé porte un ton 1. Ils identifient mieux le *doute* (DOUB) et l'*intimité* si le ton final



de l'énoncé est un ton 2. La *politesse* (POLI) est bien reconnue lorsque l'énoncé se termine par un ton 3.

En général, l'énoncé se terminant par un ton 2 reçoit le meilleur score moyen d'identification pour tous les affects sociaux aussi bien par les sujets français que par les sujets vietnamiens.

Lorsque le ton neutre est utilisé à la fin de l'énoncé, c'est la séquence tonale 4-4-0 qui reçoit le taux moyen de reconnaissance le plus élevé pour tous les affects sociaux.

En comparant les scores moyens de reconnaissance des affects sociaux obtenus pour chaque énoncé où seul le ton initial est différent, soit tous les énoncés de 4 syllabes, nous observons peu de différences entre ces énoncés à la fois chez les sujets français et chez les sujets vietnamiens.

#### 3.2.2.5. Confusion perçue des affects sociaux

Dans cette partie, afin de mieux comprendre les difficultés rencontrées par les sujets pour bien identifier les affects sociaux chinois, nous poursuivons notre travail par l'analyse des confusions entre ces affects sociaux observées chez les auditeurs français et vietnamiens.

L'analyse de la matrice de confusion présentant les confusions entre affects sociaux chez les sujets français est résumée dans la figure 48, et celle des sujets vietnamiens est représentée par la figure 49. Seules les confusions qui affichent des scores deux fois supérieurs au seuil du hasard sont représentées sur ces figures.

Dans les réponses données par les auditeurs français, la majorité des reports importants (ceux cinq fois supérieurs au hasard) se fait vers la *déclaration* (DECL). L'*ironie* (IRON), la *confiance* (CONF) et l'*évidence* (OBVI) ont été particulièrement confondues avec la *déclaration*. La confusion de l'*irritation* (IRRI) vers l'*autorité* (AUTH) et celle de la *déception* (DISA) vers la *résignation* (RESI) sont aussi observées. La *question* (QUES) ne fait pas de report vers d'autres affects sociaux, mais elle attire principalement les confusions des trois expressions d'exclamation de *surprise* et du *doute* (DOUB). Au contraire, l'exclamation de *surprise négative* n'attire aucun autre affect social, mais est particulièrement confondue avec les deux autres exclamations de *surprise* (NEU-S, NEG-S et POS-S), le *doute*, la *question* et l'*irritation*. A l'exception de la *question*, les trois

expressions de *surprise* sont également confondues avec le *doute*. Le maternel (IDS) n'est confondu qu'avec la *séduction* (SEDU), ainsi que la *résignation* (RESI) avec la *déception* (DISA). L'*admiration* (ADMI) ne montre pas de confusions particulières avec les autres affects sociaux.

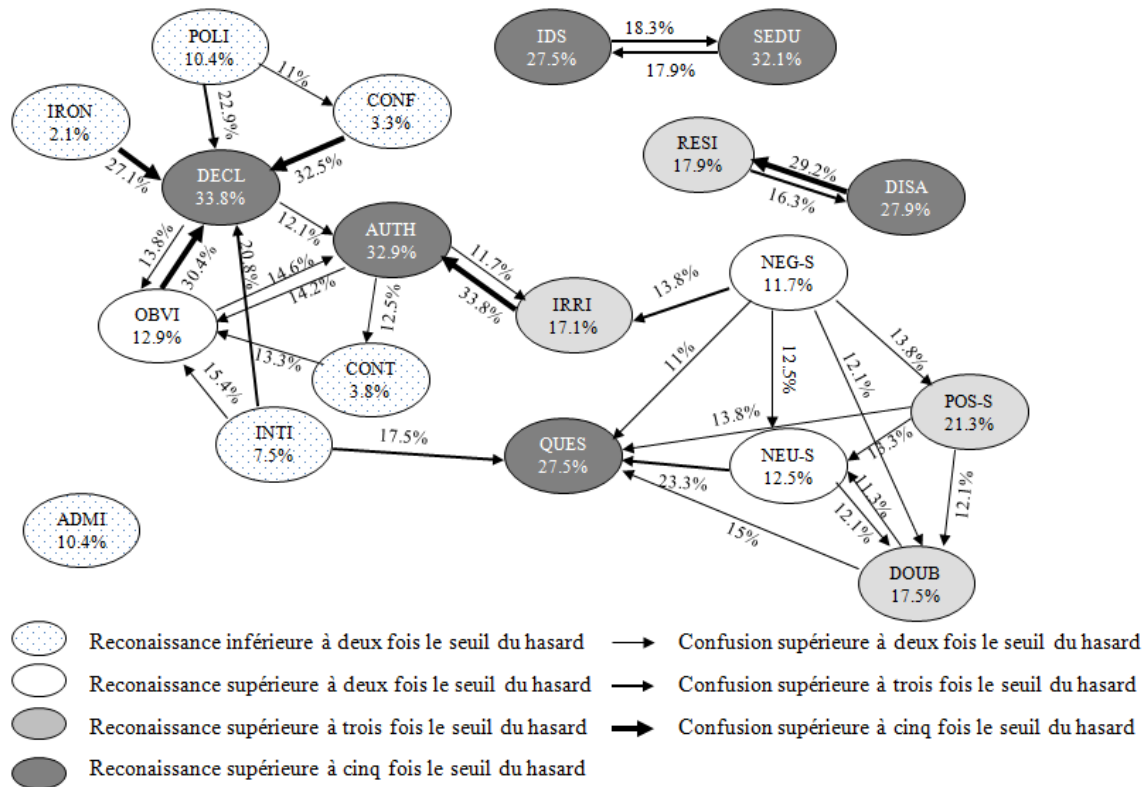


Figure 48: Confusions perçues des 19 affects sociaux chinois pour 15 sujets français : les pourcentages d'identification sont indiqués dans les ovales et les taux de confusions sur les flèches. Seules les confusions qui affichent des scores supérieurs à deux fois le seuil du hasard sont rapportées.

Pour les auditeurs vietnamiens, la plupart des reports se font aussi vers la *déclaration* (DECL), notamment la *confiance* (CONF) et la *politesse* (POLI). La *déclaration* est parfois confondue avec la *confiance* (CONF). Comme observés chez les sujets français, des liens étroits entre les attitudes de *question* (QUES), de *surprise* (NEU-S, NEG-S et POS-S) et de *doute* (DOUB) sont également mis en avant dans les réponses données par les sujets vietnamiens. Cependant, la *question* n'est pas particulièrement confondue avec d'autres affects sociaux, elle est effectivement la cible des transferts des trois expressions de *surprise* et du *doute*. L'*irritation* (IRRI) et l'*autorité* (AUTH) sont confondues, tout comme la *résignation* (RESI) et la *déception* (DISA). Les transferts de l'*irritation* vers l'*autorité* et de la *résignation* vers la *déception* se montrent majoritaires par rapport aux transferts

inverses. Comme pour les auditeurs français, l'*admiration* (ADMI) ne montre pas de confusions particulières avec les autres affects sociaux chez les auditeurs vietnamiens.

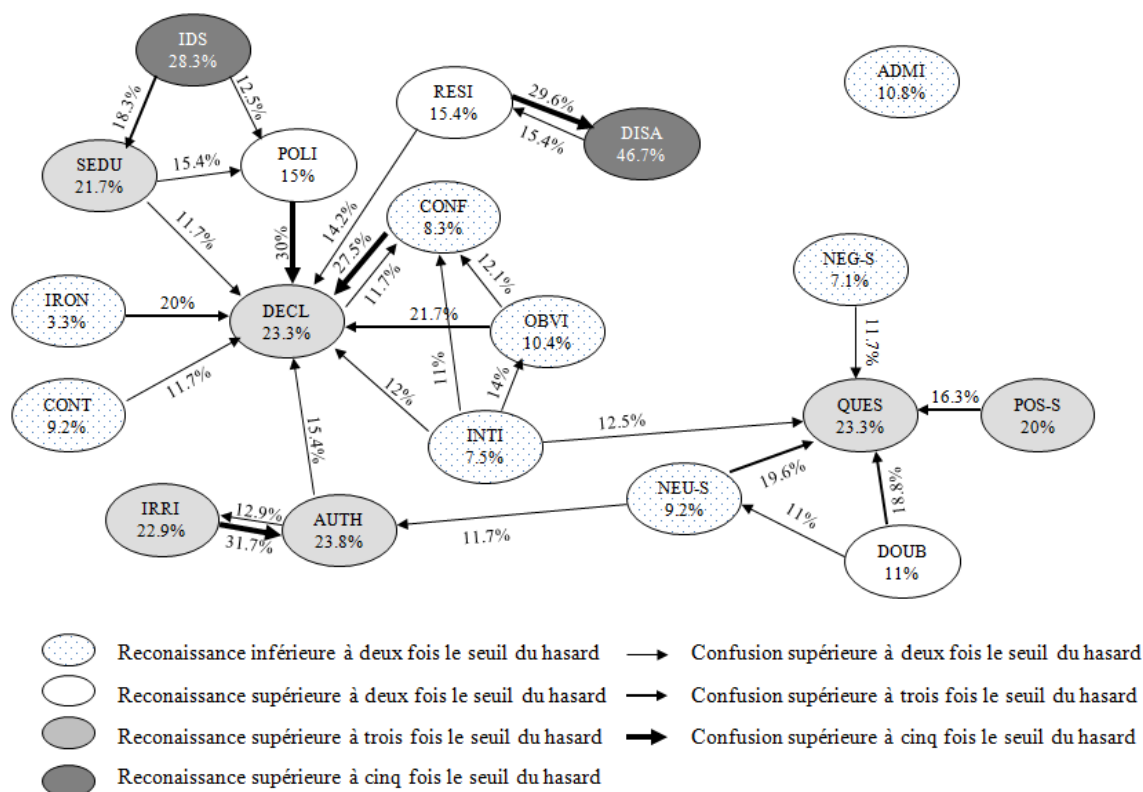


Figure 49: Confusions perçues des 19 affects sociaux chinois pour 15 sujets vietnamiens: les pourcentages d'identification sont indiqués dans les ovales et les taux de confusions sur les flèches. Seules les confusions qui affichent des scores supérieurs à deux fois le seuil du hasard sont rapportées.

Pour pouvoir comparer avec les résultats de la perception intra-culturelle des mêmes affects sociaux en chinois mandarin (voir partie 2.3), nous avons observé l'attractivité de ces affects sociaux en calculant la somme de scores de confusion provenant d'autres affects sociaux vers un affect social donné. La figure 50 représente l'attractivité de chaque affect social respectivement pour les sujets chinois, français et vietnamiens. D'après nos données, pour tous les auditeurs, la *déclaration* est l'attitude la plus « attractive », - la plus choisie comme réponse dans l'expérience perceptive. Ce fort score montre un choix plus fréquent de l'étiquette *déclaration* par le sujet, lorsqu'il n'a pas pu identifier un affect social.

Cette tendance fait écho aux comportements perceptifs des sujets brésiliens, anglais, vietnamiens et japonais pour les attitudes exprimées dans leur langue maternelle (de Moraes, 2010 ; Diaferia, 2002 ; Mac, 2010 ; Shochi, 2009). Se distinguant des sujets chinois natifs, les sujets français et vietnamiens montrent une préférence d'un deuxième

choix après la réponse *déclaration*, qui est l'*évidence* pour les sujets français et la *question* pour les sujets vietnamiens. L'*ironie* n'est pas reconnue par les auditeurs vietnamiens et n'attire pas non plus les autres affects sociaux.

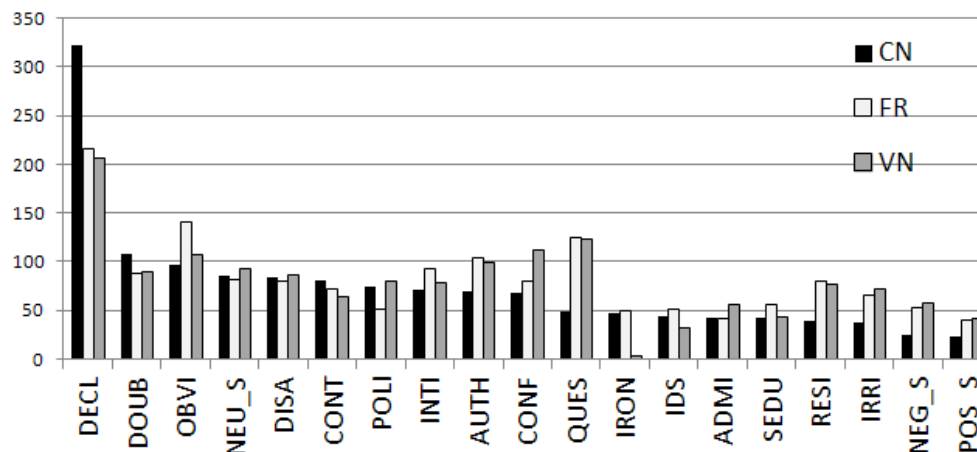


Figure 50: L'attractivité des 19 affects sociaux chinois – pourcentages cumulatifs de confusions provenant d'autres affects sociaux vers un affect social donné, pour les sujets chinois (CN), français (FR) et vietnamiens (VN).

### 3.2.2.6. Classement des affects sociaux

Afin d'extraire les tendances générales du regroupement des affects sociaux pour les deux groupes de sujets (français et vietnamiens), un algorithme de classification hiérarchique a été appliqué à la matrice de confusion, en prenant en compte l'ensemble de confusions perceptives des sujets. Les distances entre les affects sociaux ont été calculées à partir de la corrélation entre les lignes (la valeur de  $1-r$  est utilisée comme distance, où  $r$  représente la corrélation entre deux lignes de la matrice de confusion). Les classifications des 19 affects sociaux obtenues pour les deux groupes d'auditeurs sont illustrées dans la figure 51. Le regroupement des affects sociaux observable chez les sujets français est représenté par le dendrogramme du haut et celui des sujets vietnamiens par le dendrogramme du bas.

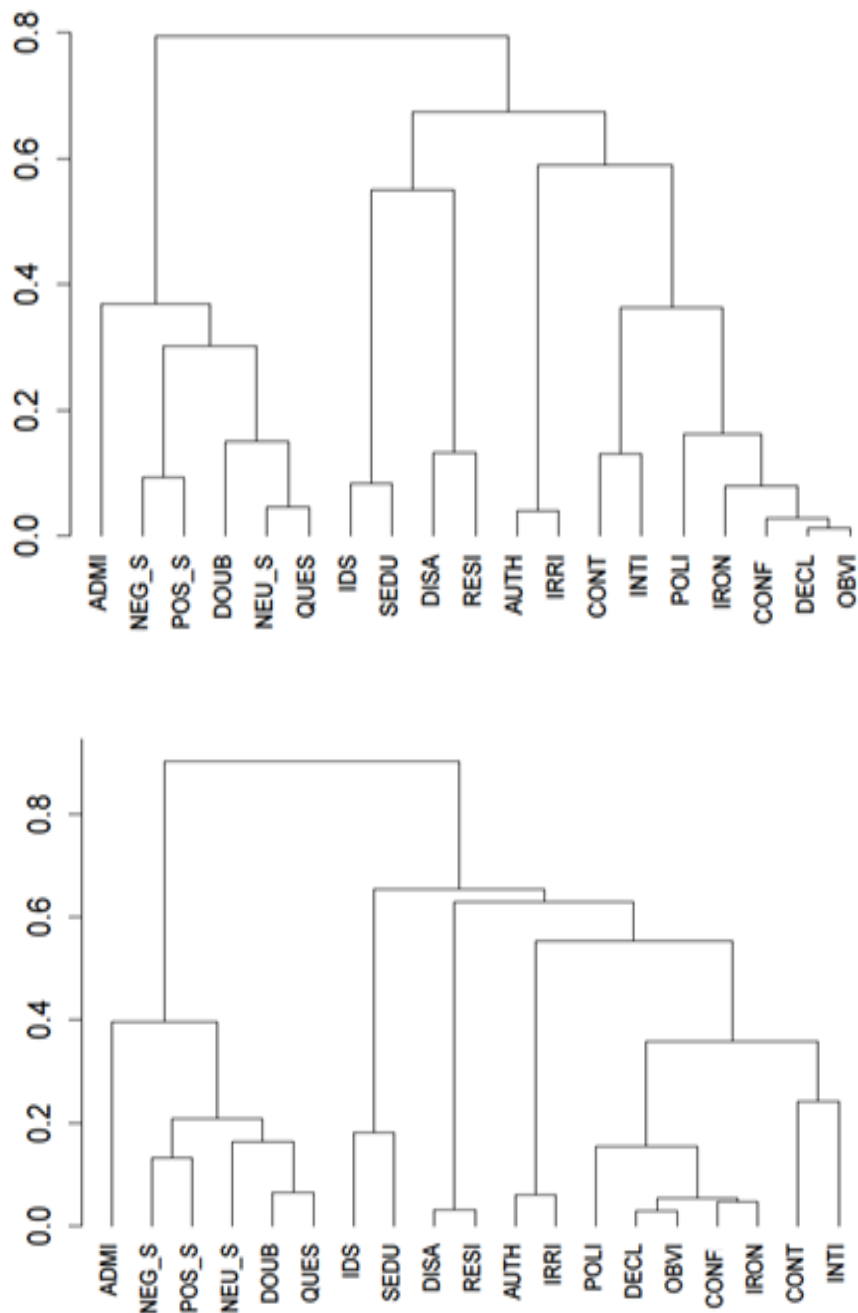


Figure 51: Classification hiérarchique des 19 affects sociaux chinois pour 15 sujets français (haut) et 15 sujets vietnamiens (bas).

En premier lieu, les deux groupes de sujets classent les affects sociaux de manière similaire. Au plus haut niveau du dendrogramme, les affects sociaux se sont divisés en deux groupes principaux : un groupe rassemble la *déclaration*, la *confiance*, l'*évidence*, la *politesse*, l'*intimité*, le *mépris*, l'*ironie*, la *résignation*, la *déception*, l'*irritation*, l'*autorité*, la *séduction* et le *maternel* ; l'autre groupe est composé des trois expressions de *surprise*, du *doute*, de la *question* et de l'*admiration*. Cette distinction correspond bien au

regroupement des mêmes affects sociaux chinois par les sujets natifs (voir partie 2.3.5.6).

Dans le cadre d'une description plus fine, les 19 affects sociaux ont été classés par les sujets français et vietnamiens en huit catégories :

- L'*admiration*, qui se caractérise par un état mental fort positif avec lequel elle se produit.
- L'exclamation de *surprise positive* et l'exclamation de *surprise négative*, qui sont des expressions inattendues, caractérisées par une activation notable.
- L'exclamation de *surprise neutre*, la *question* et le *doute*, qui sont des expressions inattendues, interrogatives et incertaines, mais moins activées que l'exclamation de *surprise positive* et de *surprise négative*.
- Le *maternel* et la *séduction*, qui expriment une proximité durable (le *maternel*) ou un souhait de proximité (la *séduction*).
- L'*autorité* et l'*irritation*, qui sont des actes de langage imposant la volonté du locuteur à son interlocuteur.
- La *déception* et la *résignation*, qui, en opposition à l'*autorité* et à l'*irritation*, reflètent des attitudes négatives, soumises et sont conceptuellement proches.
- L'*intimité* et le *mépris*, qui pourtant montrent peu de similarités que ce soit au niveau de l'état mental qu'au niveau prosodique (voir partie 2.4).
- La *déclaration*, la *confiance*, la *politesse*, l'*évidence* et l'*ironie*, qui constituent une catégorie d'affirmation sous plusieurs formes.

En comparant cette classification avec la classification des mêmes affects sociaux chinois par des sujets natifs (voir partie 2.3.5.6), nous pouvons constater que la plupart des affects sociaux sont classés de manière similaire par les sujets natifs et les sujets non-natifs. En effet, le *doute*, l'exclamation de *surprise neutre* et la *question* sont jugés acoustiquement proches par tous les sujets; le *maternel* est généralement confondu avec la *séduction*, l'*autorité* avec l'*irritation*, et la *déception* avec la *résignation*, quelle que soit la langue maternelle des auditeurs. Cependant, quelques différences sont aussi à noter essentiellement pour l'*admiration*, le *mépris*, l'*ironie* et l'*intimité*. Les sujets chinois jugent proches l'*admiration* et l'exclamation de *surprise positive*, mais les sujets non-natifs séparent l'*admiration* des autres affects sociaux. Les sujets français et vietnamiens

différencient l'exclamation de *surprise négative* et de surprise *positive* des attitudes de *doute*, de *question* et d'exclamation de *surprise neutre*, alors que les sujets natifs ne font pas cette distinction. La table 18 compare le classement des affects sociaux obtenu pour des sujets de langue maternelle chinoise et celui auditeurs de langue maternelle française et vietnamienne:

Groupe d'affects sociaux	Affects sociaux	
	Sujets Chinois	Sujets français / vietnamiens
Group 1	ADMI, POS-S	ADMI
Group 2	NEG-S, NEU-S, QUES, DOUB	NEG-S, POS-S
Group 3	IDS, SEDU	DOUB, NEU-S, QUES
Group 4	AUTH, IRRI	IDS, SEDU
Group 5	DISA, RESI	AUTH, IRRI
Group 6	IRON, CONT	DISA, RESI
Group 7	DECL, CONF, POLI, OBVI, INTI	CONT, INTI
Group 8		DECL, CONF, POLI, OBVI, IRON

Table 18: Bilan de classifications des affects sociaux chinois obtenues pour les sujets chinois, français et vietnamiens.

### 3.2.3. Conclusion

16 énoncés exprimés avec 19 affects sociaux (304 stimuli) ont été retenus pour cette étude interculturelle de la perception des affects sociaux chinois chez des sujets français et vietnamiens. Dans l'ensemble, tous les affects sociaux ont pu être relativement bien identifiés par les 30 sujets français et vietnamiens, à l'exception de l'attitude d'*ironie* dont le taux d'identification est inférieur au seuil du hasard. En comparant les comportements perceptifs des deux groupes de sujets, nous avons constaté que la *déclaration* est l'attitude la mieux reconnue par les sujets français et la *déception* celle des sujets vietnamiens. Les sujets français savent mieux identifier la *séduction*, l'*autorité*, l'*évidence*, la *déclaration*, la *question* et le *doute* que les sujets vietnamiens, alors que les sujets vietnamiens reconnaissent mieux la *déception*, l'*irritation*, la *politesse*, le *mépris* et la *confiance*. L'*ironie* n'est pas reconnue par les deux groupes de sujets (taux d'identification est inférieur au hasard).

Les résultats des analyses de variance montrent que la position et la combinaison des tons, de même que l'interaction entre les tons et les affects sociaux produisent un effet

significatif ( $p < 0,01$ ) sur l'identification de ces affects sociaux pour tous les sujets français et vietnamiens. Globalement, quand le ton des autres syllabes est fixé au ton plat, celui de la dernière syllabe de l'énoncé exerce une influence sur la reconnaissance des affects sociaux chinois. Peu de différences ont été observées lorsque le ton de la première syllabe de l'énoncé varie.

Les confusions perceptives montrent que l'*ironie*, attitude non reconnue par les sujets français et vietnamiens, est principalement confondue avec la *déclaration*. Les expressions d'exclamation de *surprise positive*, de *surprise négative* et de *surprise neutre*, ainsi que celle de *doute* sont fréquemment confondues avec la *question* par tous les sujets. En confrontant le comportement perceptif des sujets non natifs à celui des sujets natifs, nous remarquons que les deux groupes d'auditeurs jugent proches acoustiquement la *résignation* et la *déception*, tout comme l'*autorité* et l'*irritation*. D'après les résultats des analyses de classification hiérarchique basées sur les matrices de confusion, la différence majeure entre les sujets natifs et les non natifs repose essentiellement dans leur manière de catégoriser les trois expressions de *surprise*, le *doute*, la *question*, l'*admiration*, le *mépris*, l'*ironie* et l'*intimité*.

### **3.3. Perception conceptuelle et psycho-acoustique des affects sociaux**

#### **3.3.1. Introduction**

A la suite des études effectuées sur la perception intra- et interculturelle des expressions prosodiques d'affects sociaux chinois, nous avons pu observer les variations d'identification et de confusions entre ces affects sociaux. Ces observations obtenues nous ont conduits à réfléchir aux raisons potentielles des mauvaises identifications et des confusions majeures de ces affects sociaux, ainsi qu'aux décalages perceptifs entre les sujets de culture d'origine différente. Nous supposons que, mis à part les similarités prosodiques de certains affects, la problématique de la perception des affects sociaux reposerait éventuellement sur leur différence de représentation conceptuelle, liée à une distance culturelle importante.

A partir de cette hypothèse et en référence aux travaux de recherche antérieurs, nous élaborons, dans la présente section de ce chapitre, une étude spécifique aux concepts de 19 affects sociaux.



La première phase de cette étude est un test sur la distance conceptuelle entre les affects sociaux représentés par des étiquettes vernaculaires, dans le but de mesurer la proximité de ces concepts chez des sujets chinois et français. Ensuite, un test perceptif a permis de mesurer la distance psycho-acoustique entre les affects sociaux chinois, afin d'examiner la distance entre le concept et la réalisation acoustique de ces affects sociaux.

### 3.3.2. Concept des affects sociaux

Les 19 affects sociaux étudiés dans les tests perceptifs intra- et interculturels ont été retenus pour cette étude. Puisque les termes utilisés pour référer aux concepts d'affects sociaux sont spécifiques à une culture, les mots qui ont la même signification dans deux cultures différentes ne correspondraient pas forcément au même concept. Afin d'éviter l'inadéquation entre l'étiquette sémantique et le concept pour un même affect social, nous avons pris l'initiative de présenter les concepts d'affects sociaux par le biais d'une description de contextes et de situations communicatives possibles, où un affect social pourrait se produire (voir partie 2.2.1). Les étiquettes utilisées dans le test ne servent qu'à représenter les concepts auxquels elles se réfèrent. Les étiquettes d'affects sociaux français ont pu être traduites en chinois : la traduction ne présentant aucune difficulté, aucun terme n'a dû être éliminé. La table 19 présente les 19 affects sociaux étudiés dans ce chapitre :

N°	Affects sociaux	Abréviation
1	Déclaration	DECL
2	Question	QUES
3	Admiration	ADMI
4	Ironie	IRON
5	Mépris	CONT
6	Irritation	IRRI
7	Confiance	CONF
8	Résignation	RESI
9	Doute	DOUB
10	Déception	DISA
11	Evidence	OBVI
12	Exclamation de surprise neutre	NEU-S
13	Exclamation de surprise positive	POS-S

14	Exclamation de surprise négative	NEG-S
15	Politesse	POLI
16	Autorité	AUTH
17	Séduction	SEDU
18	Maternel	IDS
19	Intimité	INTI

Table 19: Liste des étiquettes des 19 affects sociaux et de leur abréviation.

### 3.3.3. Méthodologie

Romney et al. (1996, 1997, 2000) ont proposé une méthode de comparaison par paires pour mesurer les similarités et les différences interculturelles des concepts pour les termes de parenté et d'émotion. Moore et al. (2002) ont appliqué cette méthode dans leurs travaux pour les termes de couleur primaire. Ces études ont montré la pertinence et l'avantage de cette méthode afin d'évaluer non seulement les similarités et les différences entre toutes les paires d'étiquettes, mais aussi les similarités de réponses données entre des sujets de différentes langue maternelle et culture d'origine.

En référence aux études précédentes, nous avons établi un ensemble d'expérimentations semblables dans le but d'examiner la distribution et la proximité des concepts d'affects sociaux dans un cadre interculturel. Le second objectif est de comparer la distribution conceptuelle et la distribution acoustique des mêmes affects sociaux à l'intérieur de la culture chinoise afin de mettre en évidence la proximité entre le concept et la réalisation acoustique de ces affects sociaux. Ainsi, trois groupes de sujets ont participé aux tests perceptifs. Pour le test conceptuel, deux groupes de sujets chinois et français ont été sélectionnés pour juger la proximité des étiquettes d'affects sociaux. Les étiquettes sont écrites en chinois ou en français selon la langue maternelle du sujet. Pour le test psycho-acoustique, un groupe d'auditeurs chinois a été sélectionné pour estimer la distance acoustique des stimuli d'affects sociaux. Une échelle à 9 degrés permettait aux sujets d'évaluer le degré de proximité entre les affects sociaux (étiquettes ou stimuli acoustique).

### 3.3.4. Mesure de la proximité conceptuelle des affects sociaux du chinois et du français

#### 3.3.4.1. Introduction

Inspirés par les travaux sur la relation entre la représentation sémantique et la conceptualisation du monde dans de nombreux domaines (par exemple Brazill et al., 1995, Romney et al., 1998, Moore et al., 1999), nous avons l'intention d'examiner ici dans quelle mesure les concepts d'affects sociaux se ressemblent et s'éloignent aussi bien à l'intérieur d'une culture qu'entre deux cultures différentes. La culture chinoise et la culture française sont l'objet de cette étude.

#### 3.3.4.2. Sujets

Deux groupes de sujets ont participé à l'expérience : un groupe de 60 sujets chinois, d'âge moyen de 30,2 ans, et un groupe de 60 sujets français, d'âge moyen de 36,5 ans. Dans chaque groupe, la répartition entre les sujets masculins et féminins est équilibrée. Aucun sujet n'a fait état de trouble perceptif ni cognitif.

Tous les sujets chinois sont des professeurs et des étudiants de l'Université des Etudes internationales de Xi'an et de l'Université normale du Shaanxi, à Xi'an en Chine. Ils n'ont aucune connaissance de la langue et de la culture française.

La plupart des sujets français sont chercheurs du Gipsa-lab, département parole et cognition, et du Laboratoire d'Informatique de Grenoble. Ils habitent tous dans la région grenobloise et n'ont aucune connaissance de la langue et de la culture chinoise.

#### 3.3.4.3. Protocole expérimental

Le test s'est déroulé via d'une interface informatique, réalisée avec le logiciel « LiveCode ». Les 171 paires d'étiquettes d'affects sociaux ont été présentées aux sujets sur un ordinateur dans une pièce calme. Le test débutait par une présentation globale et une description de chaque affect social illustré par des exemples de contextes possibles. Ces informations étaient retranscrites sur un document papier dans la langue maternelle du sujet (en chinois pour les sujets chinois et en français pour les sujets français). Les sujets conservaient ce document comme référence pendant le déroulement du test. Après avoir lu les consignes, les sujets s'entraînaient avec une paire d'exemples non utilisée dans le test. L'objectif de cet entraînement est de familiariser les sujets avec la

tâche à réaliser : d'abord juger la similarité entre les concepts de deux affects sociaux présentés, puis évaluer la similarité sur une échelle à 9 degrés. L'échelle ne comporte pas de valeurs sauf deux symboles placés aux extrémités : à gauche « ≠ » signifie totalement différent et à droite « = » signifie totalement identique. Toutes les paires d'étiquettes ont été présentées aux sujets une seule fois et dans un ordre aléatoire. Les réponses données par les sujets ont été sauvegardées automatiquement sur le disque dur de l'ordinateur.

#### 3.3.4.4. Analyse des résultats

##### 3.3.4.4.1. Distribution des concepts d'affects sociaux

Dans un premier temps, une analyse de correspondance (AC) a été effectuée à partir de 2280 (19 affects sociaux x 120 sujets) x 19 matrices de similarité (Moore et al., 1999; Romney et al., 2000; Moore et al., 2002). Chaque matrice présente le score de similarité entre les étiquettes des affects sociaux : si les deux étiquettes comparées sont identiques, leur score de similarité est de 10, sinon, le score est égal à 10 moins la distance mesurée dans le test. Les scores de similarité obtenus dans l'AC sont normalisés avec la méthode proposée par Kumbasar et al. (1994) : pour chaque sujet, la moyenne de toutes les valeurs est réglée à 0 et la variance est égale à la valeur singulière de chaque dimension de l'AC.

En regroupant les sujets en fonction de leur langue maternelle ou de leur genre, nous constatons l'impact de ces deux facteurs sur la distribution des concepts d'affects sociaux dans l'espace perceptif (Romney et al., 2000). Les données complètes comptent 18 dimensions, dont les trois premières expliquent à 46,1% la variance totale et indiquent les principales tendances. Les figures 52 et 53 sont des représentations graphiques dues à l'AC.

La figure 52 représente la dispersion des résultats des 120 sujets regroupés par langue maternelle pour les 19 concepts d'affects sociaux. En général, peu de différences sont observées entre les sujets chinois et français. Autrement dit, les variations de la dispersion des concepts sont principalement dues aux affects sociaux eux-mêmes, plutôt qu'à la langue maternelle des sujets.

Les affects sociaux se répartissent en deux groupes dans la première dimension qui représente 28,7% de la variance. La *question* (QUES) et l'exclamation de *surprise neutre* (NEU-S) se positionnent au centre de l'espace perceptif. La *résignation* (RESI) et l'*autorité* (AUTH) ne se dispersent pas de la même manière entre les résultats des sujets chinois et ceux des sujets français. La *séduction* (SEDU) révèle aussi quelques différences entre les deux langues maternelles, les sujets chinois l'isolent des autres affects sociaux notamment dans la troisième dimension.

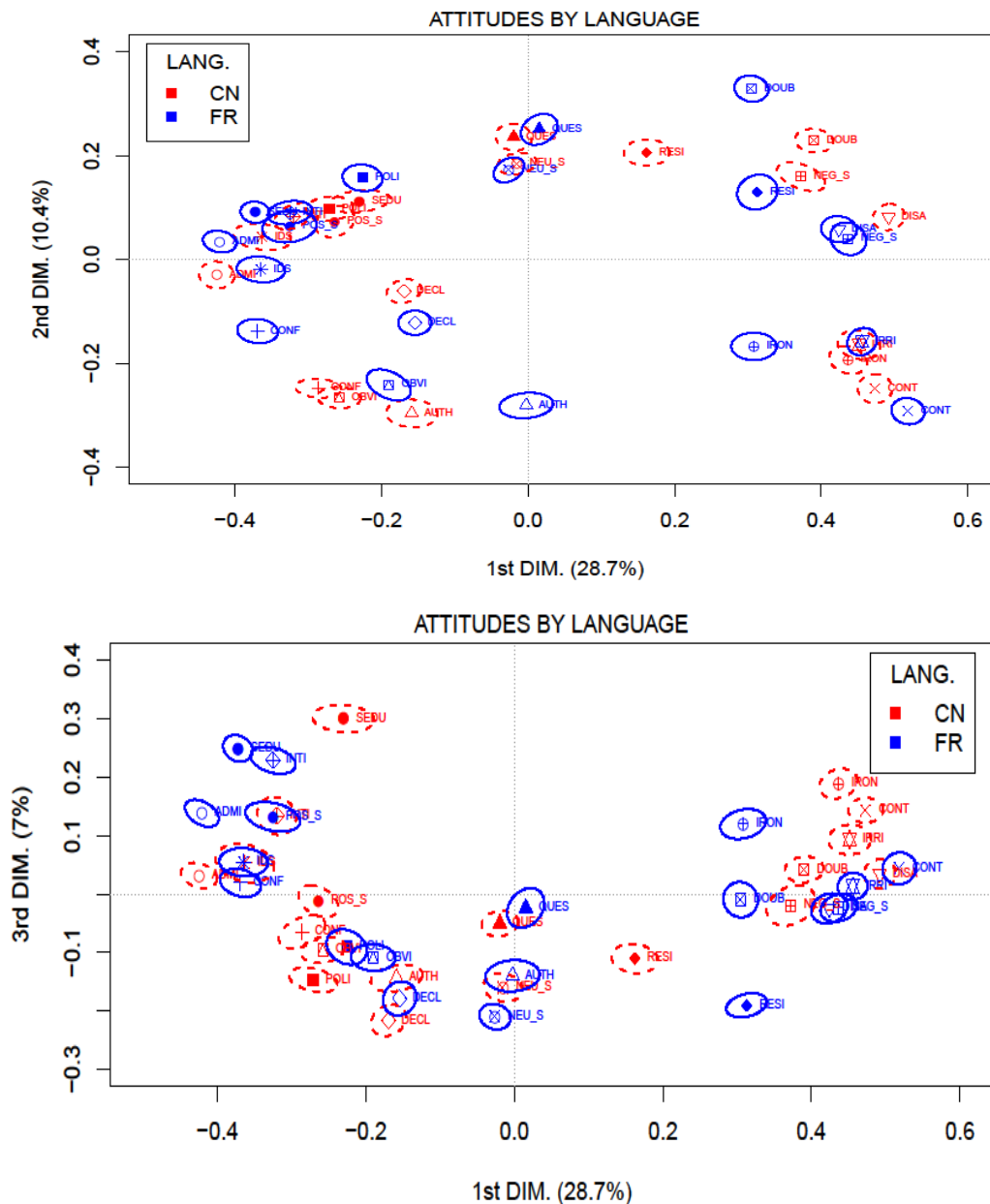


Figure 52: Distribution des 19 concepts d'affects sociaux par 120 sujets regroupés par langue maternelle dans les trois premières dimensions de l'AC. Les cercles autour de chaque concept indiquent un niveau de confiance de 97,5%.

La figure 53 illustre la distribution des affects sociaux en fonction du genre des sujets. Selon les résultats de l'analyse, le genre des sujets exerce une faible influence sur la dispersion des concepts d'affects sociaux, à l'exception de l'exclamation de *surprise positive* (POS-S) et de la *résignation* (RESI) pour la deuxième dimension.

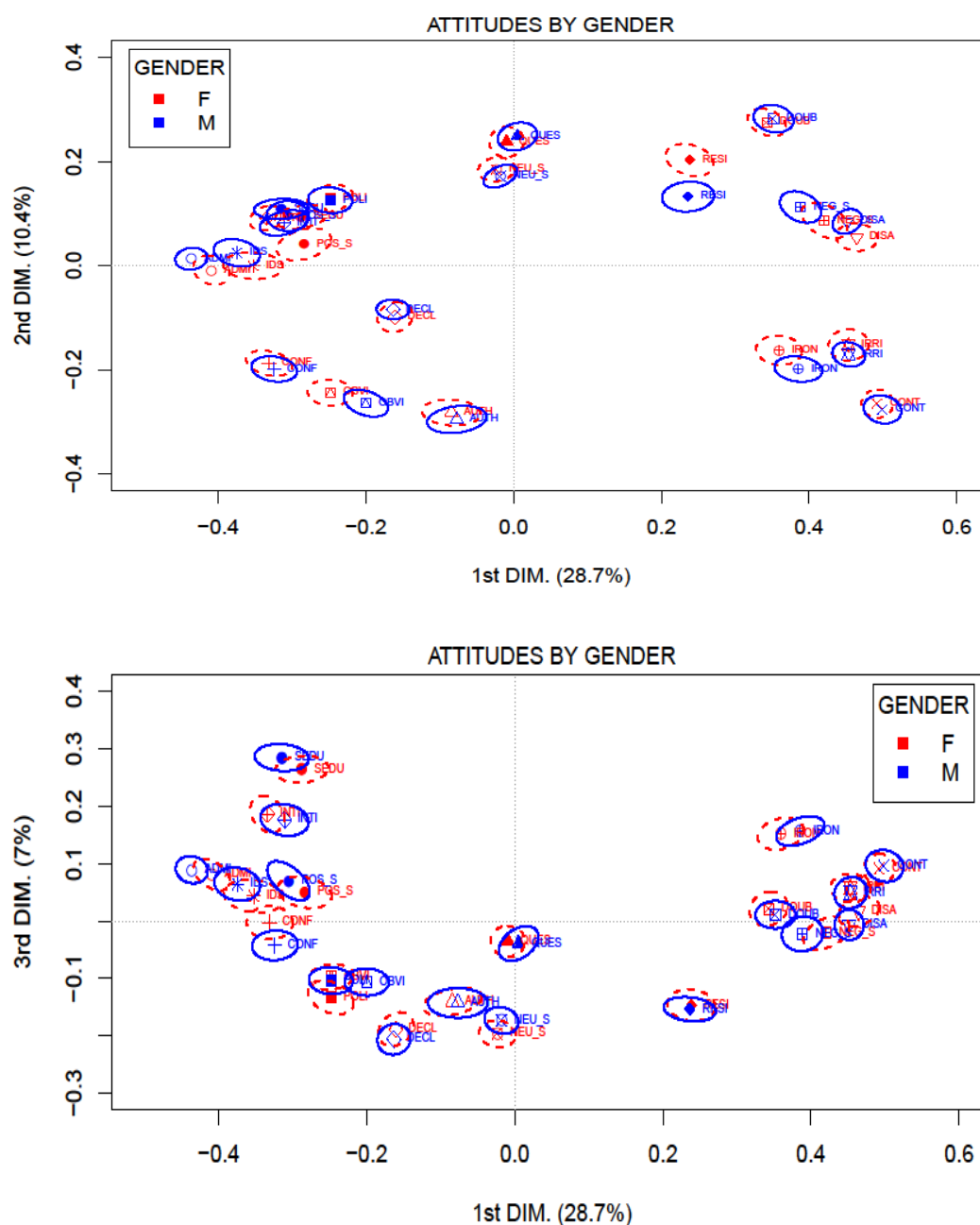


Figure 53: Distribution des 19 concepts d'affects sociaux par 120 sujets regroupés par genre dans les trois premières dimensions de l'AC. Les cercles autour de chaque concept indiquent un niveau de confiance de 97,5%.

Or, ces configurations de variations pour les trois premières dimensions ne peuvent pas représenter toutes les variations apparues dans les données où au total 18 dimensions

ont été comptabilisées. Elles ne représentent que quelques tendances générales. Afin d'apporter plus de précisions sur la similarité entre les affects sociaux, nous proposons une classification hiérarchique.

#### 3.3.4.4.2. Classement des concepts d'affects sociaux

Dans un premier temps, nous avons effectué une analyse de classification hiérarchique pour les concepts d'affects sociaux pour tous les participants. Cette analyse est basée sur les corrélations entre les affects sociaux obtenues dans l'AC présentée ci-dessus, et elle permet une observation globale du regroupement des affects sociaux. La distance euclidienne a été utilisée comme la distance entre individus; la moyenne de la distance entre deux groupes (méthode du average-linkage de R) a été utilisée pour montrer la distance entre groupes.

Nos résultats sont présentés dans le dendrogramme de la figure 54. Les concepts d'affects sociaux sont séparés en deux groupes.

Le premier groupe contient des affects sociaux « positifs », et se divise en deux sous-groupes. Le premier sous-groupe réunit l'*évidence* (OBVI), la *confiance* (CONF), la *déclaration* (DECL) et l'*autorité* (AUTH). Les trois premières attitudes se révèlent de différentes affirmations du locuteur à propos d'un fait. Elles sont marquées par une évaluation de valence relativement neutre. L'*autorité* est par contre un affect social d'imposition, qui montre le statut social ou hiérarchique supérieur du locuteur par rapport à son interlocuteur. L'autre sous-groupe rassemble le *maternel* (IDS), l'*intimité* (INTI), l'*admiration* (ADMI), l'exclamation de *surprise positive* (POS-S), la *politesse* (POLI) et la *séduction* (SEDU). Par rapport aux affects sociaux du premier sous-groupe, ceux-ci sont plus étroitement liés à la relation sociale interindividuelle et au contexte de conversation. De plus, ces expressions sont plutôt issues d'un jugement favorable sur un fait ou une personne, et ont ainsi une valence positive.

Le second groupe rassemble des affects sociaux « négatifs » : l'*ironie* (IRON), le *mépris* (CONT) et l'*irritation* (IRRI). Ils sont dus à un jugement défavorable sur un fait ou une personne, et sont donc marqués par une valence négative. Ils pourraient être considérés comme un refus du locuteur de l'opinion de son interlocuteur, mais avec un arousal différent : l'*ironie* se réalise avec un arousal plutôt faible et l'*irritation* avec un arousal plus fort. La *résignation* (RESI), la *déception* (DISA) et l'exclamation de *surprise négative*

(NEG-S) révèlent aussi certaines valeurs négatives, mais d'un arousal beaucoup plus faible que celui de l'*ironie*, du *mépris* et de l'*irritation*. La *question*, le *doute* (DOUB) et l'exclamation de *surprise neutre* (NEU-S) sont des affects causés par l'inconfort dans la croyance du fait énoncé par l'interlocuteur : ce sont des expressions interrogatives (*question*), inattendues (exclamation de *surprise neutre*) et incertaines (*doute*). Ces affects sont considérés comme affectivement neutres.

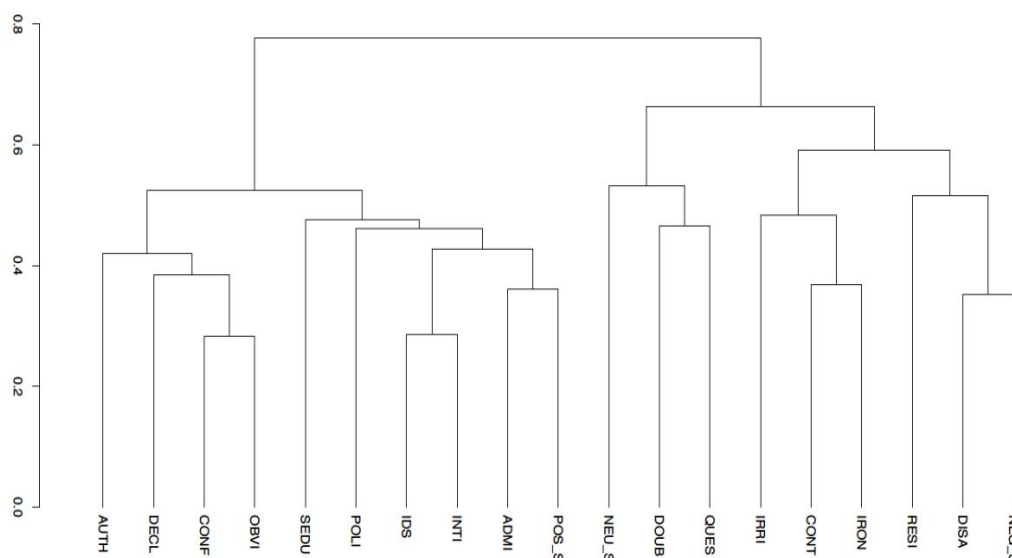


Figure 54: Classification hiérarchique des 19 affects sociaux à partir de leurs proximités conceptuelles pour les 120 sujets.

Après avoir discuté des tendances générales du regroupement des affects sociaux à partir de leurs proximités conceptuelles, nous considérons nécessaire d'examiner la spécificité de chaque langue dans la catégorisation des concepts d'affects sociaux. Par conséquent, deux analyses de classification hiérarchique distinctes ont été menées pour les sujets chinois et les sujets français.

Les dendrogrammes de la figure 55 montrent que les sujets chinois et les sujets français catégorisent les affects sociaux de manière similaire : les affects sociaux sont regroupés en deux catégories identiques, à l'exception de la *question* (QUES). Les sujets chinois jugent proches le concept de *question* (QUES) et celui de l'exclamation de *surprise neutre* (NEU-S), et les regroupent avec les affects sociaux positifs. Cependant, pour les sujets français, la *question* est proche du *doute* (DOUB), et est ainsi regroupé avec les affects sociaux négatifs.



D'autre part, la *séduction* (SEDU) et la *résignation* (RESI) sont catégorisées de manière singulière par les sujets chinois : elles sont respectivement séparées des affects sociaux positifs et des affects sociaux négatifs. Cette observation converge vers le résultat de l'AC par langue maternelle (voir figure 52).

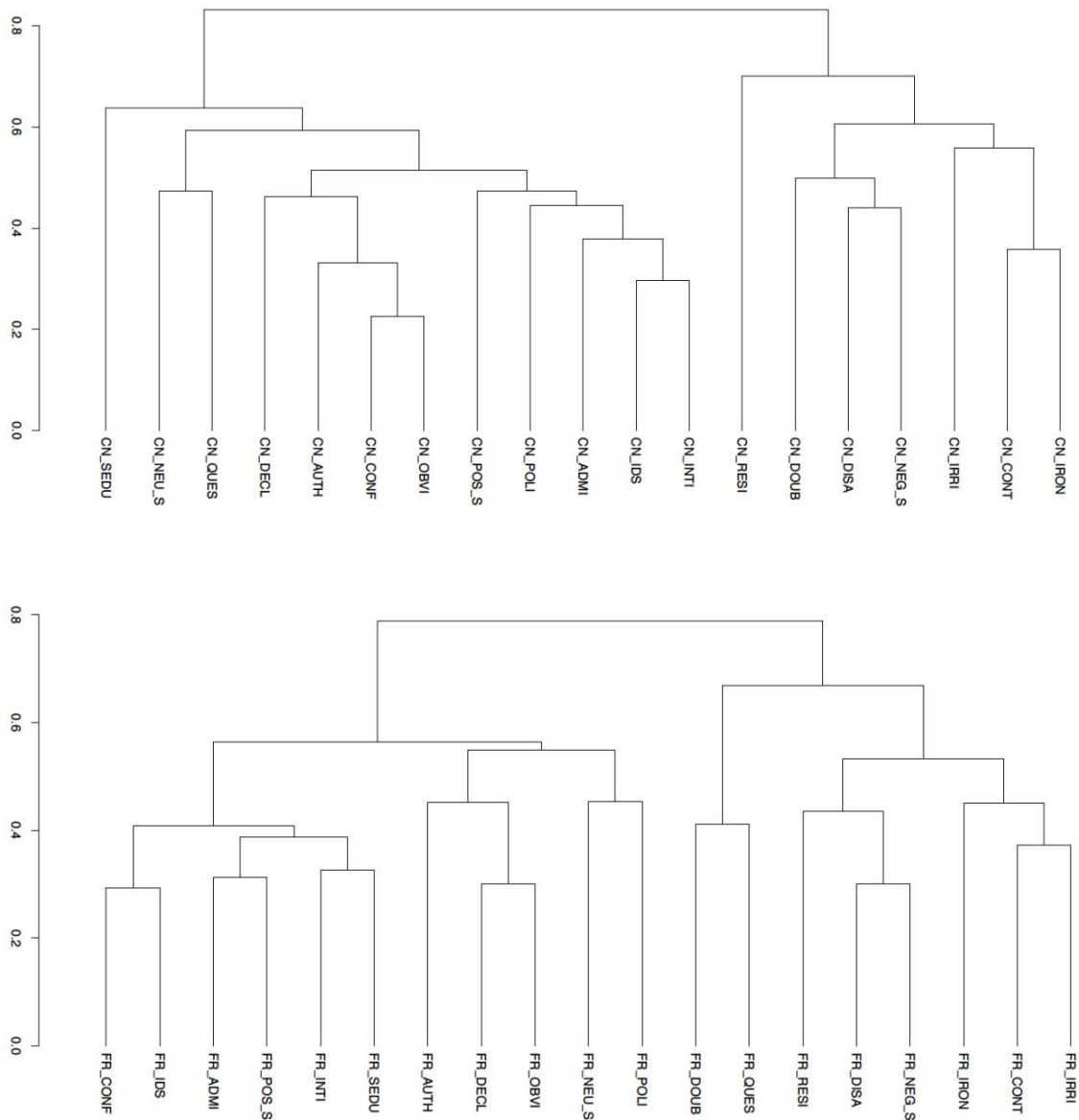


Figure 55: Classifications hiérarchiques par langue maternelle des 19 concepts d'affects sociaux à partir de leurs proximités conceptuelles : sujets chinois (haut) et sujets français (bas).

#### 3.3.4.4.3. Proportion de connaissances partagées entre les sujets

Afin de mesurer et quantifier les similarités de réponses données par les sujets, nous avons procédé à la comparaison de la forme de l'espace perceptif (Romney et al., 2000)

des sujets en fonction de leur langue et culture et de leur genre. Plus les espaces perceptifs de deux sujets sont corrélés, plus leurs connaissances des concepts d'affects sociaux sont similaires. Les valeurs sont calculées en pourcentage de connaissances partagées entre différents groupes de sujets selon des critères de langue ou de genre.

La figure 56 représente l'importance relative de différentes sources de variation contrôlées dans la perception des concepts d'affects sociaux. Le résultat montre que 64% des informations sont partagées entre les sujets sans tenir compte de leur langue et de leur genre; 3% des informations sont spécifiquement dues à la langue et culture d'origine des sujets indépendamment de leur genre ; la contribution du genre ne se limite qu'à 1,5% indépendamment du contexte linguistique et culturel des sujets. Les autres 31,5% des variations sont liées au hasard et à l'erreur d'échantillonnage.

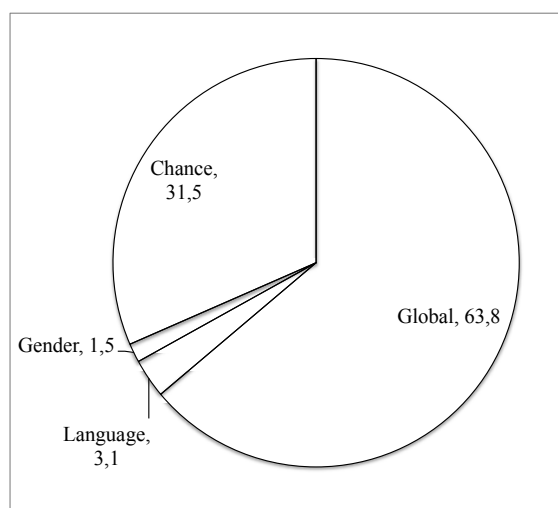


Figure 56: Contribution relative aux connaissances partagées par 120 sujets pour les 19 concepts d'affects sociaux due à ce qui est universellement partagé (Global), ce qui est spécifiquement lié à la langue et culture des sujets (Language), au genre des sujets (Gender) et à l'erreur d'échantillonnage (Chance).

#### 3.3.4.5. Conclusion

Les distances entre les 19 concepts d'affects sociaux ont été mesurées dans cette partie avec la méthode de comparaison par paires proposée par Romney et al. (1996, 1997, 2000). Tous les concepts d'affects sociaux ont été étiquetés et illustrés par une description de la situation communicative où un affect social peut être engendré. Deux groupes de sujets ont participé à ce test conceptuel: un groupe de sujets chinois et un groupe de sujets français.

Les résultats des analyses indiquent que les sujets chinois et français partagent dans une large mesure les connaissances sur les concepts d'affects sociaux. Beaucoup de similarités ont été trouvées entre les deux langues, ainsi qu'en fonction du genre. En effet la plupart des concepts sont catégorisés de manière similaire par les sujets chinois et français. Ceci implique l'existence de certains traits universaux dans la proximité conceptuelle des affects sociaux, en dépit de la distance culturelle importante entre ces deux communautés linguistiques. Néanmoins, la *question* est catégorisée différemment par les deux groupes de sujets : elle est regroupée avec les affects sociaux positifs par les sujets chinois et avec les affects sociaux négatifs par les sujets français. Au sein d'une même culture, certains affects sociaux sont isolés des autres du même groupe, ce qui suppose qu'ils ont certaines particularités en concept. C'est le cas de la *séduction* et de la *résignation* pour les sujets chinois.

Une autre observation importante porte sur la tendance générale de la catégorisation de ces concepts d'affects sociaux. Globalement, tous les concepts sont regroupés par tous les sujets en quatre groupes en fonction de leur valence et de leur arousal:

- Les affects sociaux de valence positive avec un fort arousal : l'exclamation de *surprise positive*, l'*admiration*, le *maternel*, l'*intimité*, la *politesse* et la *séduction* ;
- Les affects sociaux de valence positive avec un faible arousal : l'*évidence*, la *confiance*, la *déclaration*, sauf l'*autorité* qui a une valence positive et un arousal fort ;
- Les affects sociaux de valence négative avec un fort arousal : l'exclamation de *surprise négative*, la *déception*, l'*ironie*, le *mépris* et l'*irritation* ;
- Les affects sociaux de valence négative avec un faible arousal : la *question*, le *doute* et l'exclamation de *surprise neutre*.

Or, la positivité ou la négativité de la *déclaration* et de la *question* sont difficiles à définir, puisqu'elles possèdent deux modalités d'énonciation fondamentales qui indiquent que le locuteur n'exprime pas (ou ne veut pas, ne doit pas, ou ne peut pas exprimer) son opinion, son jugement ou son attitude. Cependant, faire référence à l'état mental et affectif du locuteur qui le conduit à formuler une assertion ou une interrogation peut permettre de mieux les expliquer.

### 3.3.5. Mesure de la proximité entre le concept et la réalisation acoustique des affects sociaux chinois

#### 3.3.5.1. Introduction

Suite à la mesure de la distribution des concepts des 19 affects sociaux, effectuée auprès de sujets chinois et français, nous explorons ici une autre mesure de la distribution de la réalisation acoustique des affects sociaux du chinois, afin de pouvoir comparer la proximité entre le concept et la prosodie de ces affects sociaux. L'objectif principal de cette étude est de vérifier si le concept et la réalisation acoustique sont étroitement corrélés pour une valeur affective donnée, et de quelle manière les différents affects sociaux s'approchent ou s'éloignent en fonction de leur distance conceptuelle et acoustique.

#### 3.3.5.2. Mesure de la distance psycho-acoustique des affects sociaux chinois

##### 3.3.5.2.1. Stimuli

Le même énoncé utilisé dans l'expérience de la perception multimodale des affects sociaux chinois a été retenu pour le test perceptif (voir partie 2.5) :

« 四天三夜 » (si4 tian1 san1 ye4, « quatre jours trois nuits » en français)

Cet énoncé a été exprimé selon les 19 affects sociaux, seulement en modalité auditive. Par conséquent, 19 stimuli audio ont été utilisés pour mesurer la distribution de la réalisation acoustique des affects sociaux chinois.

##### 3.3.5.2.2. Sujets

Un groupe de 32 auditeurs chinois (16 hommes et 16 femmes), d'âge moyen de 21 ans, a participé à cette expérience. Ils sont étudiants à l'Université des Etudes Internationales de Xi'an. Aucun n'est allé à l'étranger et n'a signalé de trouble auditif.

##### 3.3.5.2.3. Protocole expérimental

Le protocole expérimental utilisé pour le test perceptif est le même que celui utilisé pour la mesure de la distance conceptuelle des concepts d'affects sociaux, présenté dans la

partie 3.3.4.3. Cependant, les étiquettes sémantiques des 19 affects sociaux chinois sont remplacées par des stimuli audio (voir supra).

### 3.3.5.3. Comparaison entre la proximité conceptuelle et la proximité acoustique des affects sociaux chinois

Les réponses données par les 60 sujets chinois dans le test conceptuel et par les 32 sujets dans le test psycho-acoustique sont réunies dans une matrice afin de mesurer la distribution des affects sociaux représentés respectivement par des étiquettes sémantiques et par la prosodie. Cette opération est réalisée au travers d'une analyse de correspondance, suivie d'une classification hiérarchique dédiée à examiner la proximité entre le concept et la réalisation acoustique de ces affects sociaux.

#### 3.3.5.3.1. Distribution des concepts et des réalisations acoustiques des affects sociaux

Dans un premier temps, une analyse de correspondance (AC) a été effectuée à partir de 1748 (19 affects sociaux x 92 sujets) x 19 matrices de similarités (Moore et al., 1999; Romney et al., 2000; Moore et al., 2002). Chaque matrice présente le score de similarité entre les étiquettes des affects sociaux : si les deux étiquettes comparées sont identiques, leur score de similarité est de 10, sinon, ce sera 10 moins la distance mesurée dans le test. Les scores de similarité obtenus dans l'AC sont normalisés avec la méthode proposée par Kumbasar et al. (1994) : pour chaque sujet, la moyenne de toutes les valeurs est réglée à 0 et la variance est égale à la valeur singulière de chaque dimension de l'AC.

18 dimensions sont obtenues pour les données complètes, mais les deux premières expliquent 34% de la variance totale et mettent en avant les distinctions les plus importantes entre les affects sociaux présentés selon les deux modalités conceptuelles et audio. Le résultat de l'AC est illustré par la figure 57. En général, les distributions des concepts d'affects sociaux et de leur représentation acoustique montrent des tendances semblables. La majorité des affects sociaux sont regroupés de la même manière, que ce soit conceptuellement ou acoustiquement : la *séduction* (SEDU), la *déclaration* (DECL), l'*évidence* (OBVI), la *confiance* (CONF) et l'*autorité* (AUTH) sont réunies ; l'*admiration* (ADMI) et l'*intimité* (INTI) se rapprochent l'une de l'autre ; le regroupement du *mépris*

(CONT), de l'ironie (IRON) et de l'irritation (IRRI) est aussi observable, ainsi que celui de l'exclamation de *surprise neutre* (NEU-S) et de l'exclamation de *surprise positive* (POS-S).

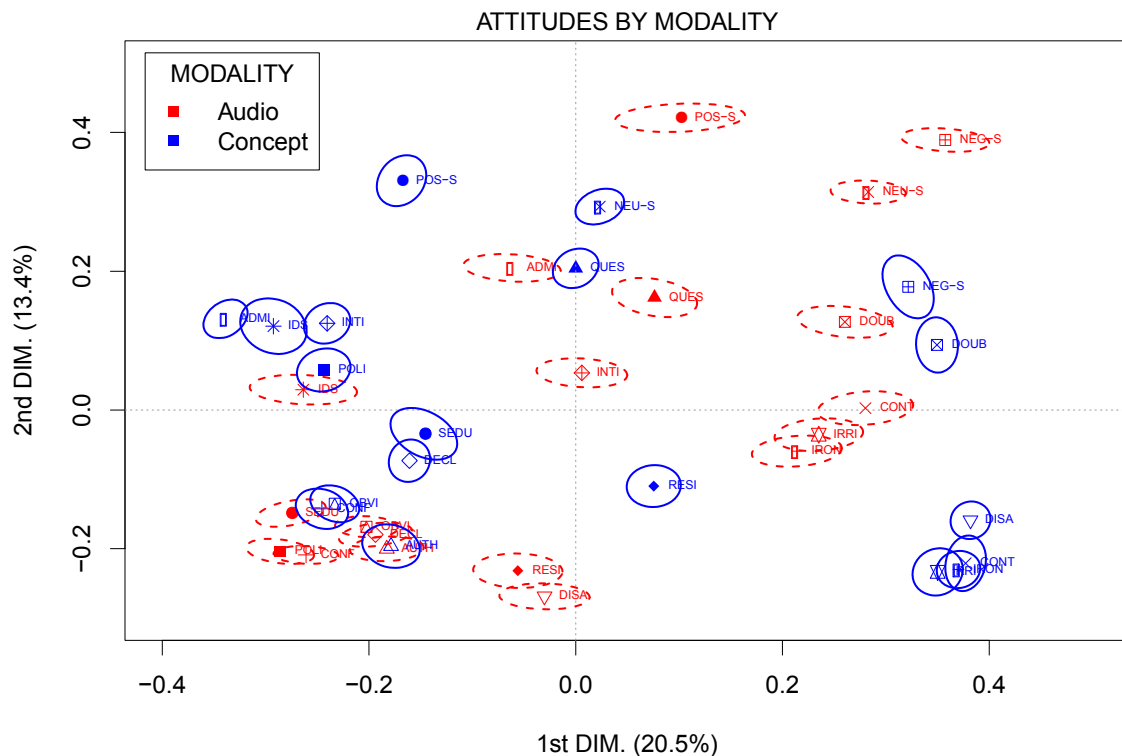


Figure 57: Distribution des 19 affects sociaux chinois présentés en concept et en audio sur les deux premières dimensions de l'AC. Les étiquettes en bleu correspondent aux concepts d'affects sociaux et les étiquettes en rouge aux stimuli auditifs; les cercles autour de chaque affect social indiquent un niveau de confiance de 97,5%.

Afin d'observer l'effet potentiel du genre sur la distribution des affects sociaux, nous avons regroupé les données selon le genre du sujet et effectué une autre AC. La dispersion obtenue est représentée par la figure 58. Le genre ne montre aucun impact majeur sur la distribution des affects sociaux que ce soit pour les stimuli conceptuels ou audio.

Cependant, puisque 18 dimensions apparaissent au total dans les données, ces configurations de variations dans les deux premières dimensions ne représentent pas toutes les variations obtenues, mais seulement quelques tendances générales. Afin d'apporter plus de précisions sur la similarité entre les affects sociaux, nous avons effectué une classification hiérarchique à partir des résultats de l'AC.



Un des groupes contient la *déclaration* (DECL), la *confiance* (CONF), l'*évidence* (OBVI), la *politesse* (POLI), l'*autorité* (AUTH), la *séduction* (SEDU), l'*intimité* (INTI), l'*admiration* (ADMI), le *maternel* (IDS), la *résignation* (RESI) et la *déception* (DISA), alors que l'autre groupe est composé du *doute* (DOUB), du *mépris* (CONT), de l'*ironie* (IRON), de l'*irritation* (IRRI), de la *question* (QUES) et des expressions de *surprise*. Chaque groupe se divise ensuite en deux. Dans le premier groupe, on peut observer une séparation entre les expressions de soumission que sont la *résignation* et la *déception* avec les expressions de *déclaration*, de *confiance*, d'*évidence*, de *politesse*, d'*autorité*, de *séduction*, d'*intimité*, d'*admiration* et de *maternel*. Pour le second groupe, la distinction observée répartit d'une part les expressions de *question* et de *surprise* et d'autre part les expressions de refus, c'est-à-dire le *doute*, le *mépris*, l'*ironie* et l'*irritation*.

Une autre observation intéressante porte sur le décalage entre le concept et la réalisation prosodique d'une même valeur affective : la *déclaration*, l'*évidence* et l'*autorité* se montrent plus proches l'une de l'autre en prosodie qu'en concept ; la *confiance* est plus proche de l'*évidence* en concept mais plus proche de la *politesse* en prosodie. Pour la *déception* et la *résignation*, leur similarité prosodique semble plus importante que leur similarité conceptuelle, ainsi que pour l'exclamation de *surprise neutre* et de *surprise négative*. Un phénomène particulier est aussi à noter pour la *séduction* : au niveau du concept, elle a été séparée des attitudes d'assertion (par exemple la *déclaration*, la *confiance*) et des autres affects étroitement liés à la relation interindividuelle et au contexte communicatif (par exemple le *maternel*, l'*intimité* et la *politesse*) ; cependant au niveau de la prosodie, elle se montre proche de la *confiance* et de la *politesse*. Ces observations suggèrent que ces affects sociaux sont mieux discriminés à travers leur concept que par leur prosodie.

D'autre part, quelques affects sociaux révèlent une proximité primitive entre leur concept et leur réalisation acoustique. La *question* (QUES), l'exclamation de *surprise positive* (POS-S), l'*intimité* (INTI), le *maternel* (IDS), l'*admiration* (ADMI), et des attitudes de refus de *doute* (DOUB), de *mépris* (CONT), d'*ironie* (IRON) et d'*irritation* (IRRI) constituent ce groupe d'affects sociaux.

La figure 60 représente la classification des affects sociaux en fonction de leur concept et de leur réalisation acoustique, ainsi qu'en fonction du genre des participants. Le résultat montre peu de différences entre les deux genres, mis à part pour la perception



acoustique de la *résignation*(RESI) : les hommes l'ont considérée plus proche de la *déception* (RESI) en prosodie, tandis que les femmes ont associé sa réalisation prosodique et son concept.

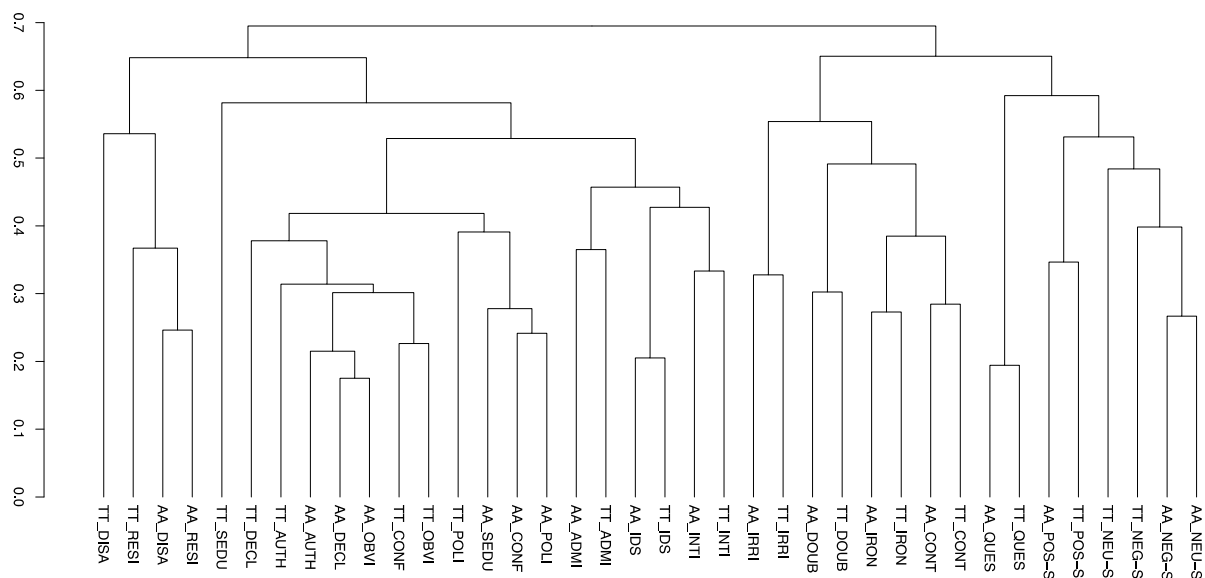


Figure 59: Dendrogramme de la classification hiérarchique des 19 affects sociaux chinois à partir de leurs proximités conceptuelles et psycho-acoustiques pour les 92 sujets natifs.

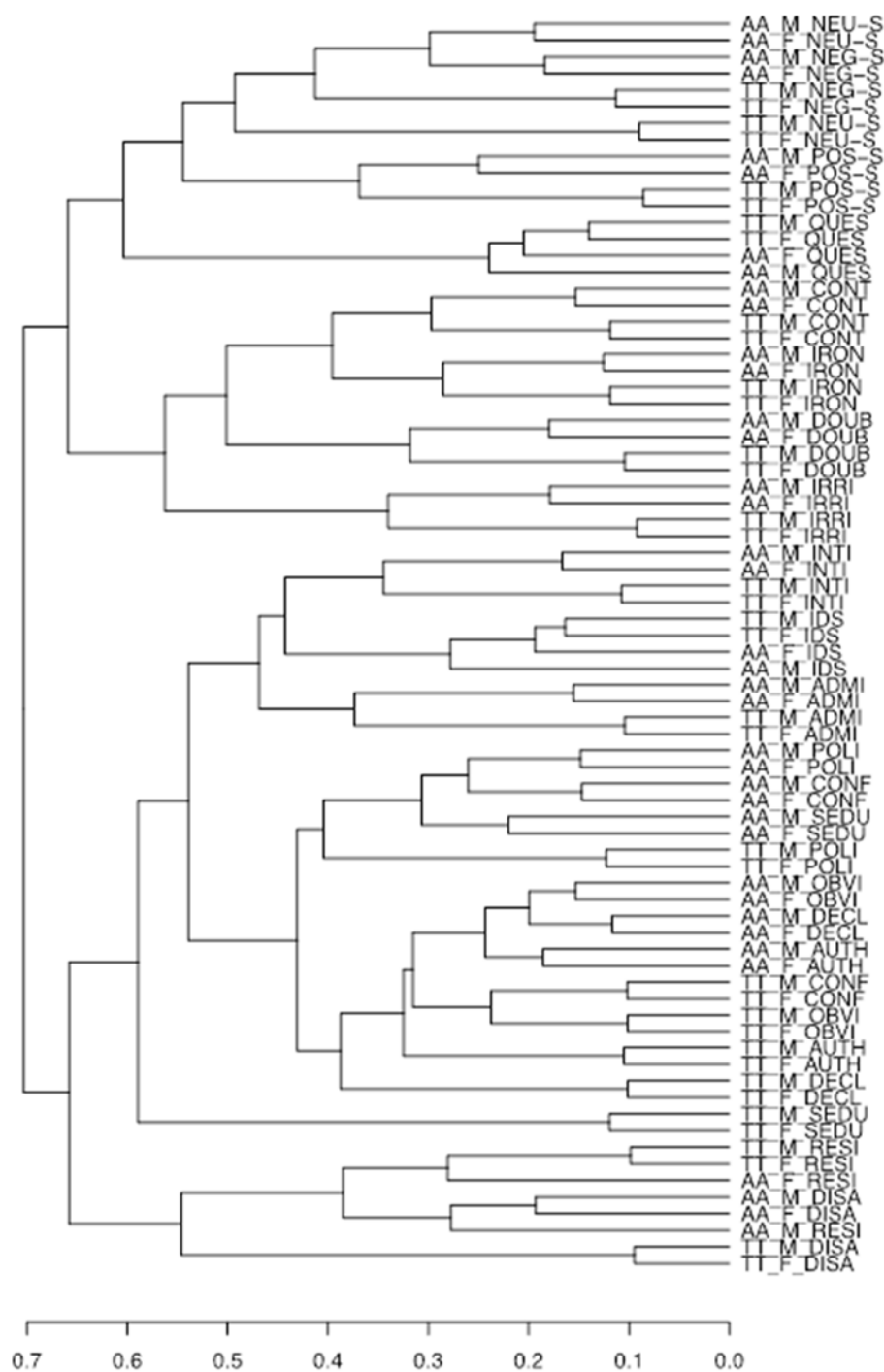


Figure 60: Dendrogramme de la classification hiérarchique des 19 affects sociaux chinois à partir de leurs proximités conceptuelles et psycho-acoustiques pour les 92 sujets natifs, en fonction du genre. AA représente les stimuli auditifs, TT les concepts, H les sujets masculins et F les sujets féminins.

### 3.3.5.3.3. Proportion de connaissances partagées entre les sujets

Afin de mesurer et de quantifier les similarités entre les réponses données par les sujets, ainsi que l'effet de la modalité de présentation des affects sociaux (concept et entrée

acoustique) et celui du genre (homme et femme), nous avons comparé la forme de l'espace perceptif des sujets en fonction du type de stimuli et du genre. Plus les espaces perceptifs de deux sujets sont corrélés, plus leurs connaissances similaires tendent à être similaires pour les données concernées.

La figure 61 illustre l'importance relative des différentes sources de variation contrôlées pendant la tâche perceptive des affects sociaux. Les sujets partagent 53% des connaissances quelles que soient la modalité de représentation des affects sociaux et le genre des sujets. En ne considérant que l'importance de la modalité de stimuli, et sans tenir compte du genre des sujets, ce pourcentage de connaissances partagées augmente à 62% pour la tâche portant sur les concepts des affects sociaux et à 64% pour celle sur leurs réalisations acoustiques. La modalité de présentation du stimulus contribue à environ 9% des connaissances partagées alors que le genre influence ce taux à seulement 2%. Le reste de 36% de variations sont liées au hasard et à l'erreur d'échantillonnage.

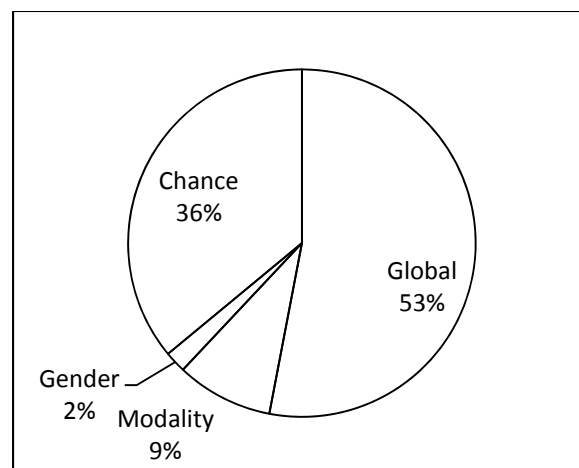


Figure 61: Contribution relative aux connaissances partagées par 92 sujets natifs pour les 19 affects sociaux chinois apportée par ce qui universellement partagé (Global), ce qui est lié à la modalité de représentation des affects sociaux (Modality), au genre des sujets (Gender) et à l'erreur d'échantillonnage (Chance).

#### 3.3.5.4. Conclusion

Les distances entre les réalisations acoustiques de 19 affects sociaux chinois ont été mesurées ici à l'aide de la méthode de comparaison par paires proposée par Romney et ses collègues (Romney et al. 1996, 1997, 2000). Cette mesure permet de comparer la

distribution acoustique des affects sociaux chinois avec leur distribution conceptuelle obtenue dans le test précédent.

Dans un premier temps, une analyse de correspondance (AC) a été réalisée sur l'ensemble des données des tests audio et conceptuels. Les résultats obtenus ont mis en avant quelques tendances générales du rapprochement entre les affects sociaux dans l'espace perceptif.

Par la suite, une analyse de classification hiérarchique basée sur les distances obtenues dans l'AC a apporté plus de précisions à ces tendances. Dans l'ensemble, tous les affects sociaux (concepts et audio) sont divisés en deux groupes principaux. Le premier groupe contient la *déclaration*, l'*autorité*, l'*évidence*, la *confiance*, la *politesse*, la *séduction*, l'*admiration*, le *maternel*, l'*intimité*, ainsi que les attitudes de soumission que sont la *résignation* et la *déception*. L'autre ensemble regroupe la *question*, l'exclamation de *surprise positive*, *négative* et *neutre*, ainsi que les attitudes de refus (le *doute*, l'*ironie*, le *mépris* et l'*irritation*). De plus, la majorité des affects sociaux de ce second groupe montre une forte corrélation entre leur concept et leur réalisation prosodique, à l'exception des expressions d'exclamation de *surprise neutre* et de *surprise négative*. En revanche, la plupart des affects sociaux du premier groupe sont marqués par un décalage entre les deux modalités de présentation, à l'exception de l'*admiration*, du *maternel* et de l'*intimité*. En effet, les perceptions prosodiques de la *déception* et de la *résignation* sont principalement corrélées entre elles, ainsi que celles de la *déclaration*, de l'*évidence* et de l'*autorité* et celles de la *politesse*, de la *confiance* et de la *séduction*, alors qu'elles ne sont pas autant similaires conceptuellement. Nous supposons que la ressemblance de ces affects sociaux est plus forte acoustiquement que conceptuellement.

Une autre analyse objective concernant les connaissances partagées entre les sujets a indiqué que 53% de connaissances sur les affects sociaux sont partagées entre les sujets sans tenir compte de la nature des stimuli (concepts ou entrée acoustique) ou du genre des sujets. Effectivement, seulement peu de variations de réponses étaient liées à la modalité (9%) et au genre des sujets (2%).

### 3.4. Conclusion du chapitre

Dans le chapitre 2, nous avons évalué la pertinence du corpus des affects sociaux chinois et nous avons également examiné l'effet de deux modalités (audio vs. vidéo), et celui de leur synergie sur la reconnaissance des expressions socio-affectives en chinois, via deux expériences perceptives auprès de sujets natifs. Dans le présent chapitre, nous nous sommes concentrés sur la perception auditive et le traitement conceptuel des mêmes affects sociaux chinois dans une étude interculturelle.

Pour l'objectif décrit ci-dessus, nous avons effectué deux tests différents :

Le premier test perceptif vise à examiner le comportement perceptif des sujets français pour les 19 affects sociaux chinois et l'effet potentiel du ton sur la reconnaissance des affects sociaux. Puisqu'ils sont considérés culturellement et linguistiquement plus proches des chinois que les français, nous avons sélectionné des auditeurs vietnamiens pour constituer un groupe de référence. 304 stimuli, à savoir 16 énoncés exprimés avec 19 affects sociaux, ont été sélectionnés pour ce test perceptif.

Au niveau du taux d'identification des affects sociaux, le résultat de l'analyse de variance ne montre pas de différence significative entre les sujets français et vietnamiens. Par conséquent, l'avantage culturel ou linguistique des sujets vietnamiens par rapport aux sujets français ne facilite pas leur reconnaissance des affects sociaux chinois, et les tons chinois seraient tout autant problématiques pour les auditeurs vietnamiens que pour les auditeurs français. Par ailleurs, l'interaction entre le ton et les affects sociaux est significative pour les deux groupes de sujets, impliquant un effet du ton sur la reconnaissance de ces affects sociaux. En observant le taux d'identification de chaque énoncé (chaque combinaison de tons) pour chaque affect social, nous avons constaté que la variation du ton final de l'énoncé influence plus la reconnaissance des affects sociaux que celle du ton initial, lorsque les autres syllabes sont fixées au ton plat (ton 1).

Au niveau des confusions perçues des affects sociaux, pour tous les sujets, bien des affects sociaux sont confondus avec la *déclaration*, ce qui correspond au comportement des sujets natifs : la *déclaration* est l'attitude la plus « attractive », c'est-à-dire la plus choisie comme réponse dans l'expérience perceptive. L'*irritation* et l'*autorité* sont essentiellement confondus l'une avec l'autre, ainsi que la *déception* et la *résignation*. La *question* n'est pas particulièrement confondue avec d'autres affects sociaux, mais elle est

effectivement la cible des transferts des trois expressions de *surprise* et du *doute*. L'*admiration* ne montre pas de confusions particulières avec les autres affects sociaux.

En matière de classement des affects sociaux, au plus haut niveau de différenciation, les affects sociaux sont divisés en deux mêmes catégories par les sujets français, vietnamiens et chinois. Dans le cadre d'une description plus fine, les sujets français et vietnamiens classent les affects sociaux en huit catégories, alors que les sujets chinois les classent en sept catégories. La différence majeure de classification des affects sociaux entre les sujets natifs et les sujets non-natifs repose essentiellement sur l'*admiration*, le *mépris*, l'*ironie*, l'*intimité*, ainsi que sur l'exclamation de *surprise positive* et de *surprise négative*.

La seconde section de ce chapitre est consacrée à l'étude de la distance conceptuelle et acoustique des affects sociaux, ainsi qu'à la corrélation entre le concept et la réalisation acoustique d'un affect social donné. Ainsi, deux tests perceptifs ont permis d'estimer les distances conceptuelles et les distances acoustiques des affects sociaux dans le but de comparer dans un premier temps la distance conceptuelle des affects sociaux chez des sujets chinois et chez des sujets français, puis dans un second temps, confronter cette distance conceptuelle à la distance acoustique au sein d'une même langue, à savoir ici le chinois.

A partir des résultats du test sur les concepts d'affects sociaux, l'analyse de correspondance a indiqué que les variations de dispersion des affects sociaux sont plus liées aux différences entre les affects sociaux eux-mêmes qu'à la langue et culture des sujets. Le genre des sujets n'a pas un effet important sur la distribution des concepts d'affects sociaux. Ensuite, selon l'analyse de classification hiérarchique, les affects sociaux sont classés de manière similaire par les sujets chinois et français. De ce fait, la conceptualisation des affects sociaux est relativement proche dans la culture chinoise et française, en dépit de leur distance culturelle. Seule la *question* est clairement classée différemment par les deux groupes de sujets. Elle est jugée proche des affects sociaux positifs par les sujets chinois, mais proche des affects sociaux négatifs par les sujets français. La *séduction* et la *résignation*, affects sociales considérées particulières par les sujets chinois, forment des isolats sur le dendrogramme.

En général, les 19 concepts d'affects sociaux sont classés par tous les sujets (sans considérer leur langue et culture) en quatre catégories différentes en fonction de leur valence et de leur arousal :

- Les affects sociaux de valence positive et avec un fort arousal : l'exclamation de *surprise positive*, l'*admiration*, le *maternel*, l'*intimité*, la *politesse* et la *séduction* ;
- Les affects sociaux de valence positive et avec un faible arousal : l'*évidence*, la *confiance*, la *déclaration*, sauf l'*autorité* qui a une valence positive et un arousal fort ;
- Les affects sociaux de valence négative et avec un fort arousal : l'exclamation de *surprise négative*, la *déception*, l'*ironie*, le *mépris* et l'*irritation* ;
- Les affects sociaux de valence négative et avec un faible arousal : la *question*, le *doute* et l'exclamation de *surprise neutre*.

En ce qui concerne la comparaison entre la distance conceptuelle et acoustique des affects sociaux pour les sujets chinois, l'analyse de correspondance indique quelques tendances générales de rapprochement entre les affects sociaux dans l'espace perceptif. Plus de détails sont mis en évidence grâce à l'analyse de classification hiérarchique basée sur les distances obtenues dans l'AC. D'après les résultats, les affects sociaux sont regroupés en deux ensembles principaux : un groupe contient les attitudes d'assertion (la *déclaration*, la *confiance*, l'*évidence* et l'*admiration*) et de soumission (la *déception* et la *résignation*), ainsi que les affects révélant la relation interpersonnelle et le contexte interactif (la *politesse*, l'*autorité*, la *séduction*, l'*intimité* et le *maternel*) ; l'autre est composé des attitudes interrogatives (la *question*), inattendues (les expressions d'exclamation de *surprise*) et incertaines (le *doute*), ainsi que des attitudes de refus (le *mépris*, l'*ironie* et l'*irritation*). En confrontant les deux groupes, nous constatons que la plupart des affects sociaux du second groupe montrent une forte corrélation entre leur concept et leur prosodie. Inversement, la majorité du premier groupe manifeste un certain décalage et est corrélée avec un autre affect social proche acoustiquement. C'est le cas de la *déclaration*, de l'*autorité*, de la *politesse*, de l'*évidence*, de la *confiance*, de la *déception*, de la *résignation* et de la *séduction*.

Enfin, globalement, la plupart des connaissances sur les affects sociaux sont partagées par tous les sujets dans les deux tests, quel que soit le contexte culturel et linguistique des sujets, leur genre et la manière de présenter les affects sociaux.

# **Chapitre 4. Etude de l'apprentissage de la prosodie socio-affective du français et du chinois mandarin**

## **4.1. Introduction**

Dans le chapitre 3, nous avons effectué des études interculturelles sur la perception et la conceptualisation des 19 affects sociaux auprès de sujets chinois et français et vietnamiens. D'après les résultats de l'analyse des données obtenues lors d'expériences, nous avons mis en évidence les similarités et les différences entre les sujets chinois et les sujets français à la fois au niveau du comportement perceptif et du traitement conceptuel des affects sociaux. Par ailleurs, nous avons montré dans le chapitre 2 que la modalité de présentation des expressions socio-affectives a un effet significatif sur la reconnaissance de ces expressions, ainsi que l'ordre de présentation des modalités.

En prenant en compte les résultats obtenus dans les études précédentes, nous allons orienter notre recherche, au sein de ce dernier chapitre de la thèse, sur la perception multimodale des affects sociaux français et chinois par des apprenants ayant acquis certaines compétences de la langue et culture cible. Cette étude a pour objectif d'examiner l'effet potentiel de la modalité de présentation des affects sociaux, ainsi que l'effet du degré d'acquisition de la langue cible sur la perception des affects sociaux chez les apprenants étrangers.

## **4.2. Hypothèse et objectif**

Il a été montré que l'acquisition perceptive de la prosodie de la langue maternelle s'amorce dès le premier mois du nouveau-né (Mehler et al., 1978), et qu'elle est déjà ancrée chez les jeunes enfants à partir de 6 mois (Crystal, 1969, 1986; de Boysson-Bardies, 1996; Billières, 2002). Clément (1999) a montré que les compétences expressives des valeurs d'attitude seraient acquises dans l'enfance entre 7 et 11 ans (à l'étape de l'installation des compétences narratives). Un large débat est aussi ouvert concernant les processus cognitifs mis en place pour l'apprentissage de la prosodie en



langue étrangère (LE) ou en langue seconde (L2)<sup>16</sup> pour les apprenants enfants ou adultes (Piske et al., 2001; Knapp, 2006). Quoi qu'il en soit, l'apprentissage de la prosodie attitudinale en LE ou L2 constitue une grande difficulté pour des apprenants adultes, parce que, d'une part, les valeurs d'attitudes / d'affects sociaux peuvent varier d'une culture à l'autre (Fónagy, 1983; Mejvaldová, 2000; Diaferia, 2002; Shochi et al., 2007). D'autre part, en tant que l'élément linguistique le plus ancré dans son comportement vocal, la prosodie de la langue maternelle de l'apprenant reste souvent rémanente dans la production en LE ou L2 (Trouvain & Gut, 2007), et est souvent génératrice de difficultés (Pettorino et al., 2011; de Meo, 2012). Ainsi, Pettorino et al. (2012) ont montré, par une expérience de transplantation réciproque de la prosodie du maître et de l'élève, que le critère prépondérant pour identifier un accent étranger est la prosodie. Par ailleurs, rappelons que l'apprentissage d'une langue étrangère et seconde ne dépend pas seulement de la langue cible, mais également de la langue source de l'apprenant (Hirschfeld & Trouvain, 2007).

De là, avant d'entamer directement l'apprentissage de la prosodie des affects sociaux français chez des apprenants chinois, nous avons mis au clair, dans un premier temps, les détails de la prosodie socio-affective de la langue maternelle de l'apprenant aussi bien sur le plan de sa perception de la prosodie que sur le plan de sa conceptualisation des affects sociaux.

La présente étude est basée sur deux hypothèses : puisque la prosodie des affects sociaux est d'ordre culturel et linguistique, la compétence en langue cible devrait avoir un effet sur la perception des expressions d'affects sociaux dans cette langue, à savoir que plus l'apprenant est compétent dans la langue cible, mieux il perçoit les affects sociaux de cette langue. Par ailleurs, comme les expressions socio-affectives varient d'une langue et d'une culture à l'autre, les apprenants d'une autre langue n'auraient pas le même comportement perceptif que les natifs pour les mêmes affects sociaux de la langue cible.

Ainsi, l'objectif principal de ce chapitre vise, en premier lieu, à examiner l'effet de l'apprentissage sur la perception des affects sociaux de la langue cible. Pour ce faire,

---

<sup>16</sup> Ici, nous distinguons la langue étrangère (LE) et la langue seconde (L2) principalement selon l'endroit où les apprenants apprennent la langue cible : les apprenants apprennent une langue étrangère dans leur pays natif, mais apprennent une langue seconde dans le pays de la langue cible.

nous comparons le comportement perceptif des apprenants de différents niveaux (niveaux débutant et intermédiaire). En second lieu, nous confrontons la perception des apprenants en LE ou L2 avec celle des natifs pour les mêmes expressions affectives, afin d'essayer de mettre en lumière les affects sociaux problématiques pour les apprenants lors de leur perception.

D'ailleurs, en considérant que les apprenants d'une langue étrangère ou seconde ont toujours besoin de se situer dans une situation face-à-face, dynamique et interactive, nous pensons qu'il est nécessaire de savoir de quelle manière ils utilisent les informations auditives et visuelles durant la perception des affects sociaux de cette langue. Par conséquent, l'effet de la modalité de présentation des expressions socio-affectives est également examiné dans cette étude.

### **4.3. Construction du corpus audio-visuel des affects sociaux français**

Un corpus audio-visuel des affects sociaux français a été conçu et enregistré pour l'étude sur l'apprentissage de la prosodie socio-affective française par des apprenants chinois.

#### **4.3.1. Sélection des affects sociaux**

Les affects sociaux utilisés dans l'étude de la perception multimodale des affects sociaux chinois (voir partie 2.5) ont été à nouveau retenus pour cette étude. Les affects sociaux sont la *déclaration* (DECL), l'*évidence* (OBVI), le *doute* (DOUB), la *question* (QUES), l'*autorité* (AUTH), l'exclamation de *surprise neutre* (NEU-S), la *déception* (DISA), le *mépris* (CONT), l'*irritation* (IRRI), la *politesse* (POLI) et le *maternel* (IDS).

#### **4.3.2. Composition du corpus**

Un seul énoncé de quatre syllabes a été utilisé pour exprimer les onze affects sociaux en français. Cet énoncé est cohérent avec celui utilisé lors de l'expérience multimodale des affects sociaux chinois : « 四天三夜 », « si<sup>4</sup> tian<sup>1</sup> san<sup>1</sup> ye<sup>4</sup> » en pinyin et « quatre jours trois nuits » en français. Cet énoncé en français est donc syntaxiquement identique à celui en chinois : 2 syntagmes nominaux de 2 syllabes énumérés sans coordinateur. L'énoncé en français est : « un jour deux nuits ».

### 4.3.3. Enregistrement du corpus

#### 4.3.3.1. Locuteur

Une femme de langue maternelle française, spécialiste de la parole expressive, a été choisie comme locutrice de ce corpus. Elle parle un français standard sans accent.

#### 4.3.3.2. Enregistrement

L'enregistrement du corpus français s'est déroulé dans le même environnement et avec les mêmes matériels que l'enregistrement du corpus chinois (voir partie 2.2). Étant donné la difficulté de produire ces affects sociaux en dehors de tout contexte, une situation de communication a été imaginée pour chaque affect social, afin d'aider la locutrice à le produire le plus naturellement possible. Une autre locutrice native prenait en charge la projection des diapositives. Elle contrôlait en même temps, avec une chercheuse en parole expressive, la performance de la locutrice pendant le déroulement de l'enregistrement. Elles observaient la locutrice en la voyant et l'entendant par le biais d'un système vidéo, et elles se sont permis d'arrêter la locutrice quand elles trouvaient qu'un énoncé n'était pas naturel, ainsi que de le lui faire reproduire. La répétition jugée la meilleure à la fois par les surveillantes et la locutrice a été sélectionnée pour cet énoncé.

Le corpus final est composé de 33 stimuli (1 énoncé \* 11 affects sociaux), plus un énoncé audio-visuel, dédié à l'entraînement du sujet. Les signaux audiovisuels ont été par la suite découpés semi-automatiquement au niveau des énoncés à l'aide du logiciel «VirtualDub».

## **4.4. Perception multimodale des affects sociaux français par des sujets natifs**

### 4.4.1. Introduction

A la suite de l'élaboration du corpus audio-visuel des affects sociaux français, nous avons effectué une expérience perceptive multimodale, d'une part, pour évaluer perceptivement la pertinence du corpus et examiner le comportement perceptif des sujets natifs pour ces affects sociaux français ; d'autre part, pour observer les impacts de la modalité de l'expression socio-affective et de l'ordre de présentation des modalités sur la perception des sujets.

#### 4.4.2. Sujets

14 participants français natifs (4 hommes et 10 femmes), d'âge moyen de 31,1 ans, ont participé à cette expérience. Tous sont soit des étudiants soit des chercheurs du Gipsa-Lab (département parole et cognition) habitant dans l'agglomération grenobloise. Aucun n'a signalé de trouble auditif ou visuel.

Durant le test, les sujets ont été aléatoirement divisés en deux groupes. Les modalités ont été présentées à chacun des groupes selon un ordre différent: les sujets du premier groupe (7 sujets) ont passé les trois modalités dans l'ordre : « audio seul », « vidéo seul », puis « audio-vidéo » ; et les sujets du second groupe (7 sujets) les ont passées dans l'ordre : « vidéo seul », « audio seul », puis « audio-vidéo ».

#### 4.4.3. Protocole expérimental

Le test perceptif s'est déroulé via une interface informatique, réalisée à l'aide du logiciel « LiveCode », pour Windows. Le test débutait par une présentation globale et une description de chaque affect social illustré par des exemples de contextes possibles. Ces informations étaient écrites dans la langue maternelle du sujet. Les participants conservaient ce document comme référence pendant le déroulement de test. Après avoir lu les consignes et avoir effectué un entraînement, les sujets commençaient le test. Les affects sociaux ont été ensuite présentés dans chacune des trois modalités de présentation, modalités elles-mêmes présentées dans un ordre donné pour chaque groupe (voir supra). A l'intérieur de chaque condition expérimentale, chaque affect social a été présenté une seule fois, dans un ordre aléatoire, différent pour chaque sujet. Pour chaque stimulus, les sujets devaient juger l'affect social exprimé par la locutrice en choisissant une étiquette parmi les 11 proposées à l'écran. Aucune limitation de temps de réflexion avant la présentation du stimulus suivant n'était imposée. A la fin du test, les réponses données avaient été sauvegardées automatiquement sur le disque dur de l'ordinateur.

#### 4.4.4. Analyse des résultats

##### 4.4.4.1. Facteurs influant la reconnaissance des affects sociaux

Les réponses des sujets ont été d'abord codées avec deux chiffres : 0 représente une mauvaise réponse (le participant ne reconnaît pas l'affect social présenté) ; 1 représente

une bonne réponse (le participant reconnaît l'affect social présenté). Par la suite, une régression logistique a été effectuée sur ces données pour étudier l'effet de plusieurs facteurs sur la variable réponse binaire. Les facteurs fixés sont la modalité de l'expression socio-affective (audio seul, vidéo seul ou audio-vidéo), l'ordre de présentation des modalités (audio seul, vidéo seul, puis audio-vidéo ou vidéo seul, audio seul, puis audio-vidéo) et l'affect social présenté (11 affects sociaux). Le modèle de régression logistique est une approche standard pour analyser la relation entre une variable dépendante binaire et des variables explicatives, dont le grand intérêt est d'évaluer l'impact des variables explicatives sur la probabilité du risque lié à la variable dépendante (Lu & Yang, 2012).

Par ailleurs, puisque dans notre expérience un participant est interrogé plusieurs fois sur une même valeur d'affect social, des effets individuels se rajoutent éventuellement aux effets fixes détaillés ci-dessus. De ce fait, nous avons choisi d'utiliser une régression logistique mixte pour nos données. Cette méthode permet alors d'être plus précis concernant l'explication des variations de la variable réponse. La régression logistique mixte a été réalisée à l'aide de la fonction « glmer » du package « lme4 » du logiciel R.

Une procédure descendante (*backward* en anglais) a été utilisée pour sélectionner les variables significatives. Le modèle initial de la procédure comprend les 3 facteurs et leurs interactions d'ordre 2. A chaque étape, la variable dont la valeur de  $p$  de l'effet associée au test de rapport de vraisemblance est la plus grande et supérieure à 0,05 est retirée du modèle. Le processus s'arrête lorsque toutes les variables sont retirées du modèle ou lorsque la valeur de  $p$  est inférieure à 0,05. Lorsqu'un facteur est impliqué dans une interaction significative, il ne peut pas être retiré du modèle même s'il est testé comme non significatif.

La table 20 donne le résultat de la sélection descendante. La première colonne présente le nom de l'effet et les trois colonnes suivantes donnent respectivement la valeur de la statistique du test, le nombre de degrés de liberté et la probabilité du test de rapport de vraisemblance. Selon le résultat, toutes les trois interactions entre les trois facteurs fixés se révèlent significatives pour la reconnaissance, par conséquent, les facteurs fixés restent tous dans le modèle, et sont considérés comme ayant un effet significatif sur la variable réponse. Les facteurs significatifs sont l'affect social, la modalité, l'ordre de présentation des modalités, ainsi que les trois interactions entre les facteurs.

	LRT $\chi^2$	ddl	<i>p</i>
Modalité : Affect social	66,398	20	0,000
Ordre de présentation des modalités : Affect social	27,483	10	0,0022
Modalité : Ordre de présentation des modalités	7,0741	2	0,0291

Table 20: Résultat de l'analyse de régression logistique mixte pas à pas descendante : les effets présentés dans le tableau sont testés comme significatifs selon un test de rapport de vraisemblance ( $p < 0,05$ ).

#### 4.4.4.2. Effet de la modalité et de l'ordre de présentation des modalités

Afin d'éclairer l'effet de la modalité sur les choix catégoriels moyens des sujets pour les affects sociaux, nous avons effectué une analyse de comparaison multiple en calculant, pour chaque affect social, les différences de proportions de bonne réponse entre les trois modalités. L'analyse a été réalisée à l'aide de la fonction « lsmeans » du package « lsmeans » du logiciel R.

- Effet de la modalité

Selon le résultat de la comparaison multiple, le *mépris*, le *doute*, le *maternel* et l'*évidence* montrent des différences significatives en modalités de présentation en matière de reconnaissance. Pour le *mépris*, les taux de reconnaissance en « audio-vidéo » et en « vidéo seul » sont significativement plus élevés qu'en « audio seul » ( $p < 0,05$ ) ; pour le *doute*, la progression du taux de reconnaissance est significative entre « audio seul » et « audio-vidéo » ( $p < 0,05$ ) ; pour le *maternel*, les taux sont significativement plus élevés en « audio-vidéo » et en « audio seul » qu'en « vidéo seul » ( $p < 0,01$ ) ; pour l'*évidence*, le taux de reconnaissance en condition audiovisuelle est significativement supérieur à celui obtenu en condition visuelle seule ( $p < 0,05$ ).

Ce résultat est illustré par la figure 62, qui représente le taux de reconnaissance moyen obtenu pour chacun des affects sociaux dans chacune des trois modalités. En général, presque tous les affects sociaux sont reconnus par les sujets natifs dans toutes les modalités (le taux d'identification est supérieur au seuil du hasard), sauf la *question* (QUES) en « vidéo seul ».

Par rapport à l'information auditive, l'information visuelle joue un rôle majeur dans la reconnaissance de l'expression de *mépris* (CONT). A l'inverse, les sujets ont plus recours à l'information auditive qu'à l'information visuelle pour reconnaître le *maternel* (IDS). Il est intéressant de noter que la *question* n'est pas reconnue par les participants natifs en

modalité visuelle seule (0%), par contre elle est bien reconnue en modalité audio seule. Ainsi, il se peut que la reconnaissance de la *question* soit plus dépendante de l'information auditive. Pour le reste des affects sociaux : l'*autorité* (AUTH), la *déclaration* (DECL), la *déception* (DISA), le *doute* (DOUB), l'*irritation* (IRRI), l'exclamation de *surprise neutre*, l'*évidence* (OBVI), la *politesse* (POLI), l'information auditive et visuelle coopèrent et se complètent l'une avec l'autre dans l'expression.

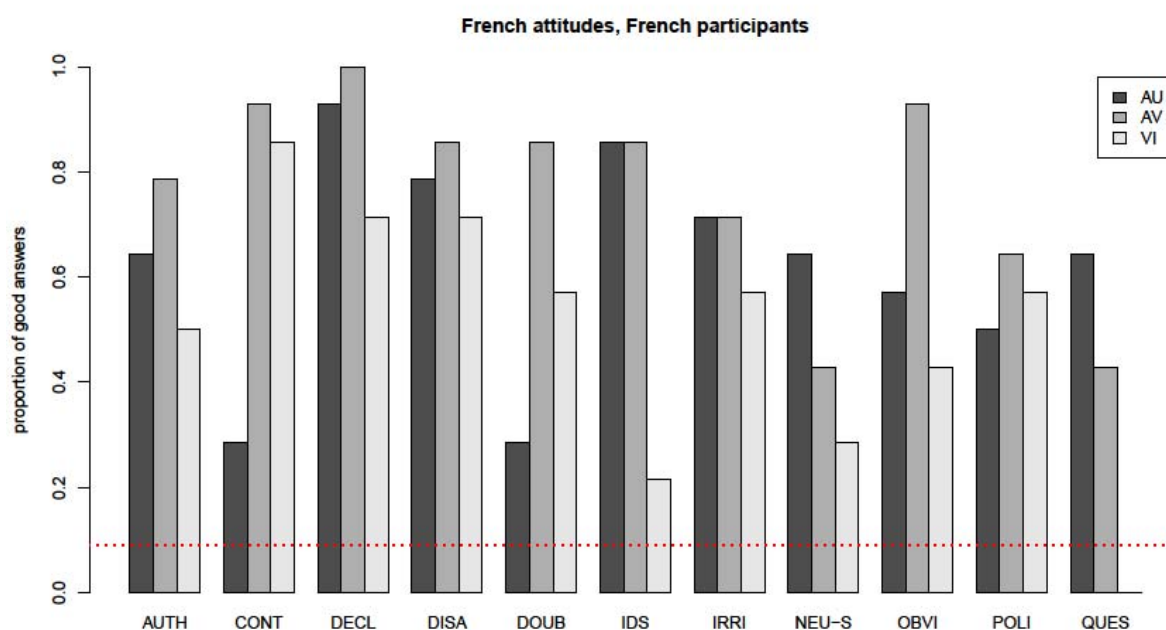


Figure 62 : Effet de la modalité sur les taux d'identification des 11 affects sociaux français perçus par des sujets français. Les affects sociaux sont présentés en abscisse et les scores de reconnaissance en ordonnée. La ligne rouge pointillée indique le seuil du hasard (9%).

- Effet de l'ordre de présentation des modalités

Selon le résultat de la comparaison multiple, l'ordre de présentation des modalités exerce un effet significatif sur la reconnaissance du *mépris*, du *doute* et de l'*irritation* ( $p < 0,05$ ).

Ce résultat est illustré par la figure 63. Globalement, les deux groupes relatifs à l'ordre de présentation des modalités ont des scores de reconnaissance similaires pour presque tous les affects sociaux. Cependant, le *mépris* (CONT), le *doute* (DOUB) et l'*irritation* (IRRI) montrent des différences entre les deux groupes de sujets : les sujets qui ont passé en premier la modalité visuelle seule identifient significativement mieux le *mépris* et l'*irritation* que les sujets qui ont passé d'abord la modalité audio seule, alors que les derniers reconnaissent significativement mieux le *doute* que les premiers.

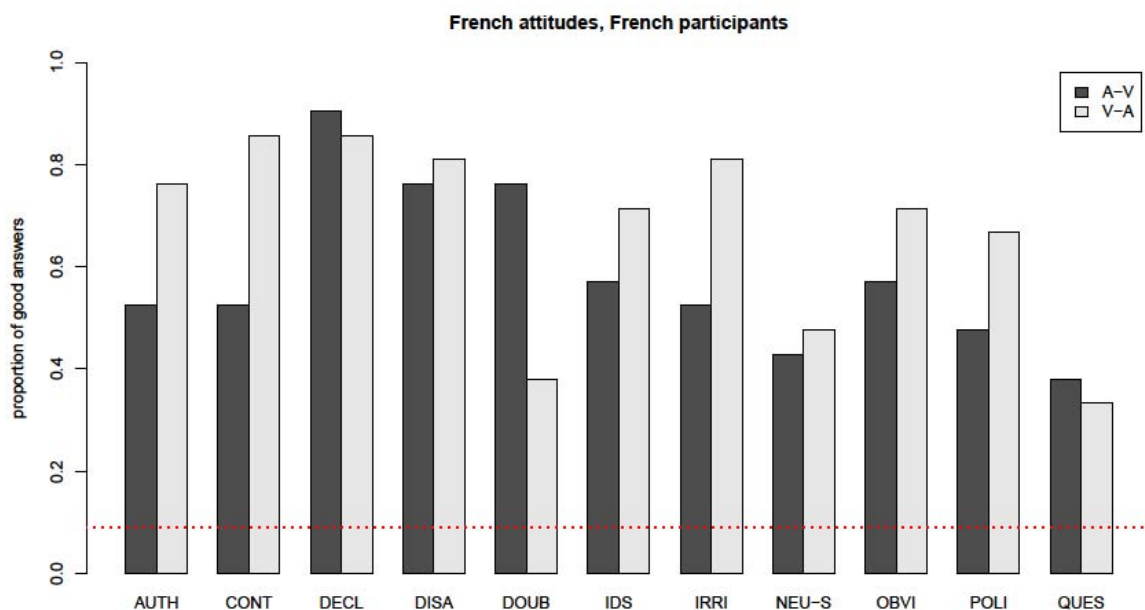


Figure 63 : Effet de l'ordre de présentation des modalités sur les taux d'identification des 11 affects sociaux français perçus par des sujets français. Les affects sociaux sont présentés en abscisse et les scores de reconnaissance en ordonnée. La ligne rouge pointillée indique le seuil du hasard (9%).

Cependant, selon le résultat de la comparaison multiple, il n'y a pas de différences significatives entre les deux ordres de présentation des modalités, pour chacune des trois modalités ( $p > 0,05$ ). Ceci est contraire au résultat de l'analyse de régression logistique. Il est possible de l'expliquer par le fait que pour certaines modalités, la proportion de bonne réponse dans les deux ordres de présentation des modalités n'est pas significativement différente, mais la différence entre les différentes modalités dans un ordre donné est significative.

#### 4.4.4.3. Confusion perçue des affects sociaux

L'analyse des confusions entre les affects sociaux a été réalisée pour les 14 sujets français, sur la base de la matrice de confusion de chaque modalité. Les résultats des deux groupes de sujets ont été mélangés. Cette analyse a été effectuée en deux étapes : la première étape consiste à discuter les confusions les plus importantes entre affects sociaux et l'orientation de ces confusions, ce qui correspond bien à la problématique de la perception de ces affects sociaux dans une modalité donnée ; la seconde étape est basée sur une analyse de classification hiérarchique de ces expressions socio-affectives dont le but est d'extraire les tendances générales du regroupement des affects sociaux en considération de l'ensemble de confusions perceptives apparues lors de l'expérience.



Lors de l'analyse de classification hiérarchique, le carré de la distance Euclidienne a été utilisé comme distance entre les observations, qui ont été alors regroupées en classes différentes selon le critère de variance minimale de Wald. En vue de déterminer le nombre optimal de groupes sur l'arbre hiérarchique, nous nous sommes référés au critère de réduction de la variance (*elbow criterion* en anglais) expliquée par chaque groupe (Soni Madhulatha, 2012). Selon ces résultats, les 11 affects sociaux peuvent être répartis en groupes qui varient selon la modalité considérée.

- Modalité « audio seul »

La figure 64 illustre les principales confusions perceptives entre les affects sociaux pour les 14 sujets natifs dans la condition audio seule, obtenues à partir de la matrice de confusion. Les confusions deux fois supérieures au hasard (9%) sont proposées.

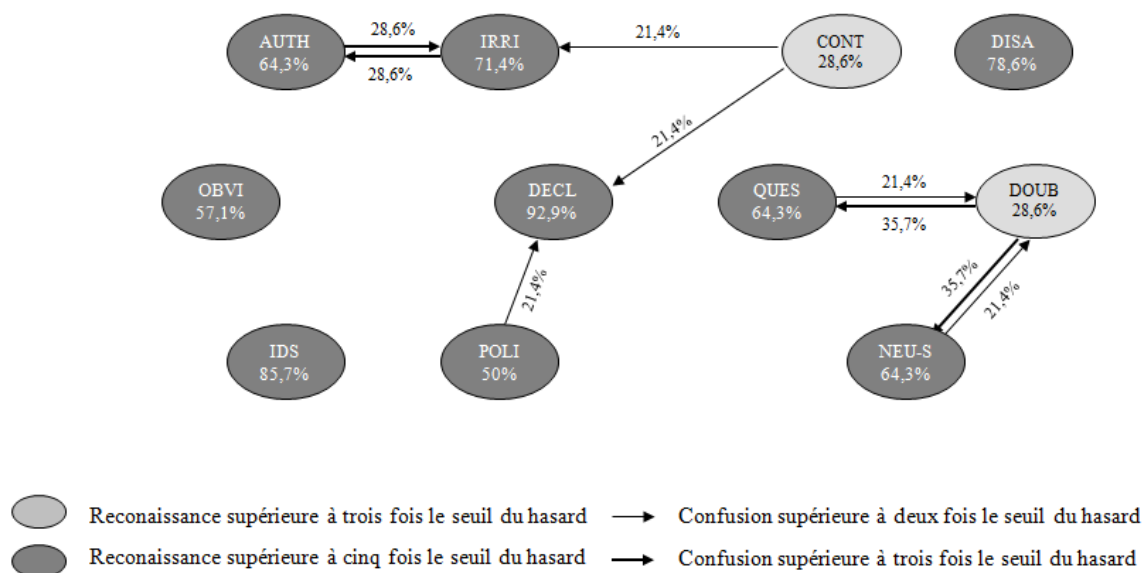


Figure 64 : Confusions des 11 affects sociaux français pour les 14 sujets français dans la modalité « audio seul ». Les pourcentages d'identification sont indiqués dans les ovales et les taux de confusions sur les flèches. Seules les confusions qui affichent des scores supérieurs à deux fois le seuil du hasard sont rapportées.

Selon le résultat, la majorité des expressions socio-affectives reçoit un score de reconnaissance supérieur à 45%. Quelques confusions perceptives sont également remarquées : la *politesse* (POLI) et le *mépris* (CONT) sont principalement confondus avec la *déclaration* (DECL) ; le *mépris* est aussi mélangé avec l'*irritation* (IRRI), qui a des confusions réciproques avec l'*autorité* (AUTH). L'exclamation de *surprise neutre* (NEU-S) et le *doute* (DOUB) montrent des confusions réciproques, ainsi que la *question* (QUES) et

le doute. La *déception* (DISA), le *maternel* (IDS) et l'*évidence* (OBVI) sont bien reconnus et ne montrent pas de confusions particulières avec les autres affects sociaux.

Par la suite, en considérant toutes les confusions entre affects sociaux, nous avons mené une analyse de classification hiérarchique afin d'identifier des classes perceptives plus larges de ces affects sociaux français pour les sujets natifs dans la condition « audio seul ».

Le dendrogramme de la figure 65 représente les résultats de cette analyse. Le graphique situé en haut à droite du dendrogramme illustre la perte d'inertie en fonction du nombre de groupes, à partir duquel on peut conclure qu'il est raisonnable de considérer 6 catégories d'affects sociaux.

Dans cette condition expérimentale, la *déclaration* (DECL), la *déception* (DISA) et le *maternel* (IDS), affects sociaux bien reconnus par les sujets, forment des isolats faute de confusions particulières avec d'autres affects sociaux. Les autres affects sociaux sont classés en trois catégories : la première catégorie contient le *mépris* (CONT), l'*évidence* (OBVI) et la *politesse* (POLI); la deuxième catégorie est composée de l'*autorité* (AUTH) et de l'*irritation* (IRRI), deux affects sociaux d'imposition; la troisième catégorie regroupe le *doute* (DOUB), l'exclamation de *surprise neutre* (NEU-S) et la *question* (QUES).

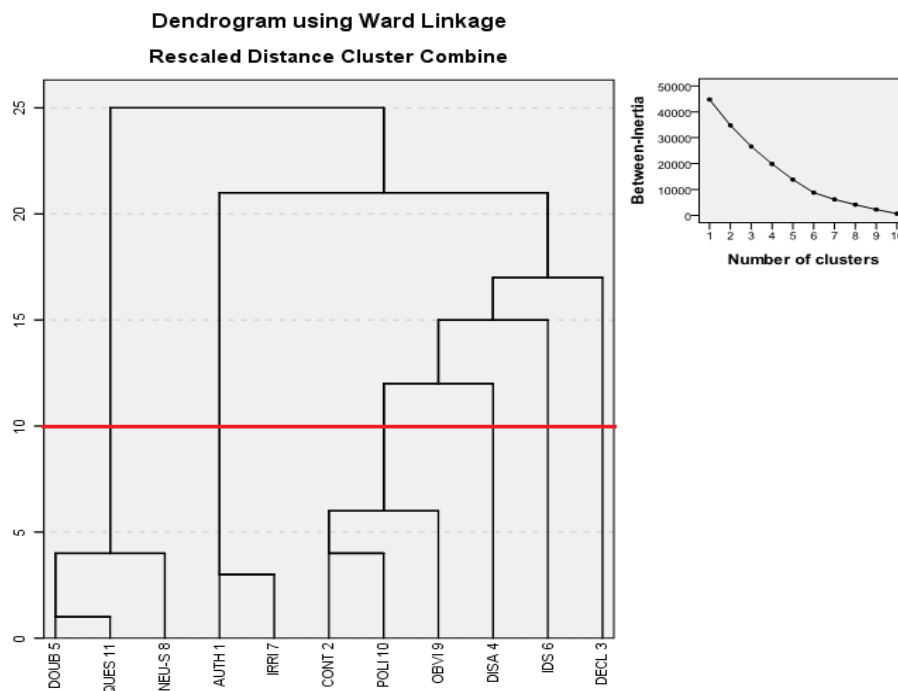


Figure 65 : Classification hiérarchique des 11 affects sociaux français pour les 14 sujets français dans la modalité « audio seul ».

- Modalité « vidéo seul »

La figure 66 représente les taux d'identification et les plus importantes confusions des affects sociaux dans la modalité visuelle seule. La *déclaration* (DECL) est bien reconnue et attire des confusions provenant du *maternel* (IDS), de la *politesse* (POLI) et de la *question* (QUES), mais elle-même ne montre pas de reports particuliers sur les autres affects sociaux. La *question* n'est pas reconnue par les sujets natifs dans cette condition expérimentale (0%) et entretient des confusions importantes avec la *déclaration* (57,1%). Le *maternel* est aussi confondu avec la *politesse* et l'*évidence* (OBVI). L'*autorité* (AUTH) et l'*évidence* sont confondues réciproquement et l'*irritation* (IRRI) est particulièrement confondue avec l'*autorité*. Les expressions de *déception* (DISA), d'exclamation de *surprise neutre* (NEU-S), de *mépris* (CONT) et de *doute* (DOUB) ne montrent pas de confusions particulières avec les autres affects sociaux.

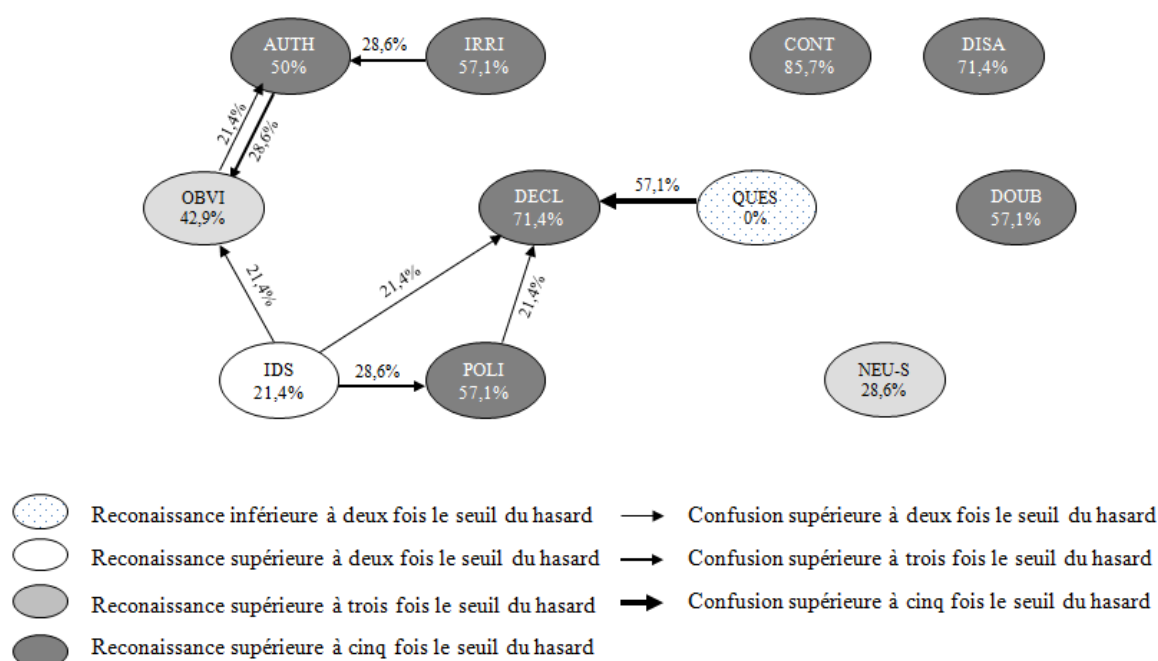


Figure 66 : Confusions des 11 affects sociaux français pour les 14 sujets français dans la modalité « vidéo seul ». Les pourcentages d'identification sont indiqués dans les ovales et les taux de confusions sur les flèches. Seules les confusions qui affichent des scores supérieurs à deux fois le seuil du hasard sont rapportées.

En ce qui concerne le regroupement des affects sociaux sur la base de leur proximité dans la modalité visuelle seule, la figure 67 illustre les résultats de classification hiérarchique. D'après la graphique de perte d'inertie en fonction du nombre de groupes, il est raisonnable de considérer 5 catégories d'affects sociaux.

La première catégorie contient la *déclaration* (DECL) et la *question* (QUES), qui sont plus particulièrement confondues sur la base des indices visuels. La deuxième catégorie regroupe la *politesse* (POLI) et le *maternel* (IDS). La troisième catégorie est composée de l'*évidence* (OBVI), de l'*autorité* (AUTH) et de l'*irritation* (IRRI). L'exclamation de *surprise neutre* (NEU-S), le *doute* (DOUB) et la *déception* (DISA) sont classés dans la quatrième catégorie à cause des confusions importantes entre eux. Le *mépris* (CONT) est bien reconnu et ne montre pas de confusions particulières avec les autres affects sociaux. Elle forme ainsi un isolat dans cette condition expérimentale.

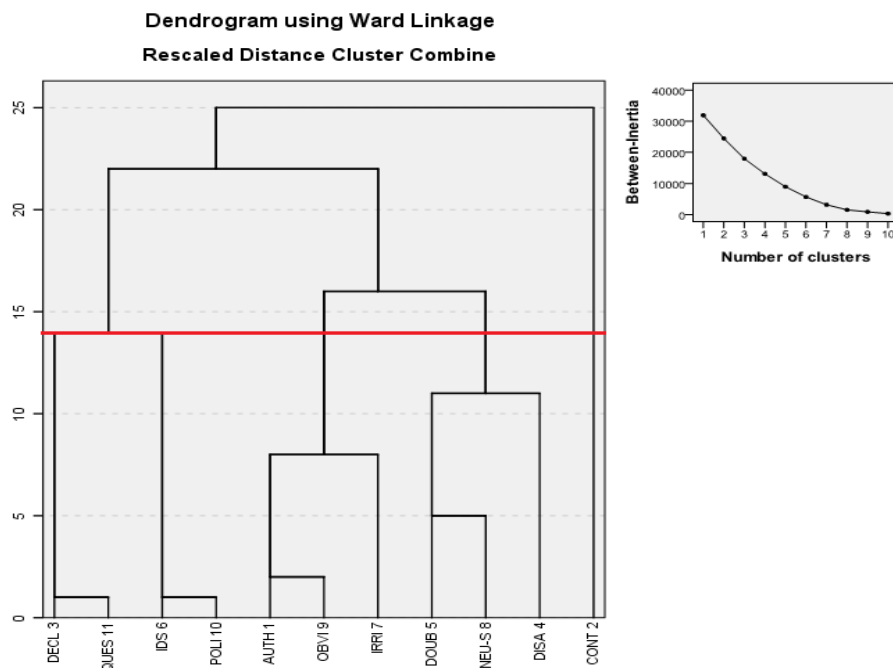


Figure 67 : Classification hiérarchique des 11 affects sociaux français pour les 14 sujets français dans la modalité « vidéo seul ».

- Modalité « audio-vidéo »

Les principales confusions entre affects sociaux pour tous les sujets dans la modalité « audio-vidéo » sont schématisées à la figure 68. Par rapport aux modalités « audio seul » et « vidéo seul », la modalité audiovisuelle véhicule plus d'informations expressives grâce à la synergie entre ces deux modalités. Par conséquent, les sujets reconnaissent mieux les affects sociaux dans cette condition expérimentale que dans les deux conditions séparées. Généralement, tous les affects sociaux sont bien reconnus dans cette modalité et la plupart d'entre eux reçoivent un score de reconnaissance cinq fois au-dessus du hasard. On observe donc naturellement moins de confusions dans

cette modalité. La *déclaration* (DECL) est reconnue sans faute (100%). Cependant, la *question* (QUES) et l'exclamation de *surprise neutre* (NEU-S) montrent des confusions réciproques, et la dernière se reporte aussi sur le *doute* (DOUB). L'*irritation* (IRRI) est parfois confondue avec l'*autorité* (AUTH), ainsi que la *politesse* (POLI) avec le *maternel* (IDS). L'*évidence* (OBVI), la *déclaration* (DECL), la *déception* (DISA) et le *mépris* (CONT) sont bien identifiés et sans confusions particulières.

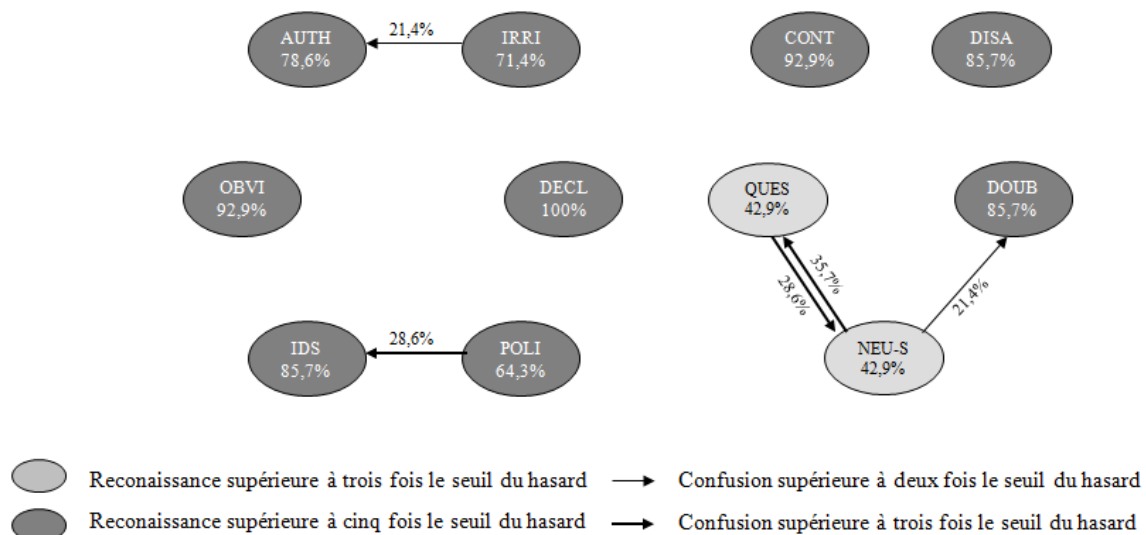


Figure 68 : Confusions des 11 affects sociaux français pour les 14 sujets français dans la modalité « audio-vidéo ». Les pourcentages d'identification sont indiqués dans les ovales et les taux de confusions sur les flèches. Seules les confusions qui affichent des scores supérieurs à deux fois le seuil du hasard sont rapportées.

La figure 69 illustre les résultats de la classification hiérarchique de ces affects sociaux sur les mêmes données dans la condition « audio-vidéo ». Les onze affects sociaux se répartissent en 7 catégories dans cette condition expérimentale.

La *déclaration* (DECL), le *mépris* (CONT), la *déception* (DISA) et l'*évidence* (OBVI), attitudes bien reconnues, forment des isolats dans cette modalité. L'exclamation de *surprise neutre* (NEU-S), la *question* (QUES) et le *doute* (DOUB) sont classés ensemble, ainsi que le *maternel* (IDS) et la *politesse* (POLI), l'*autorité* (AUTH) et l'*irritation* (IRRI).

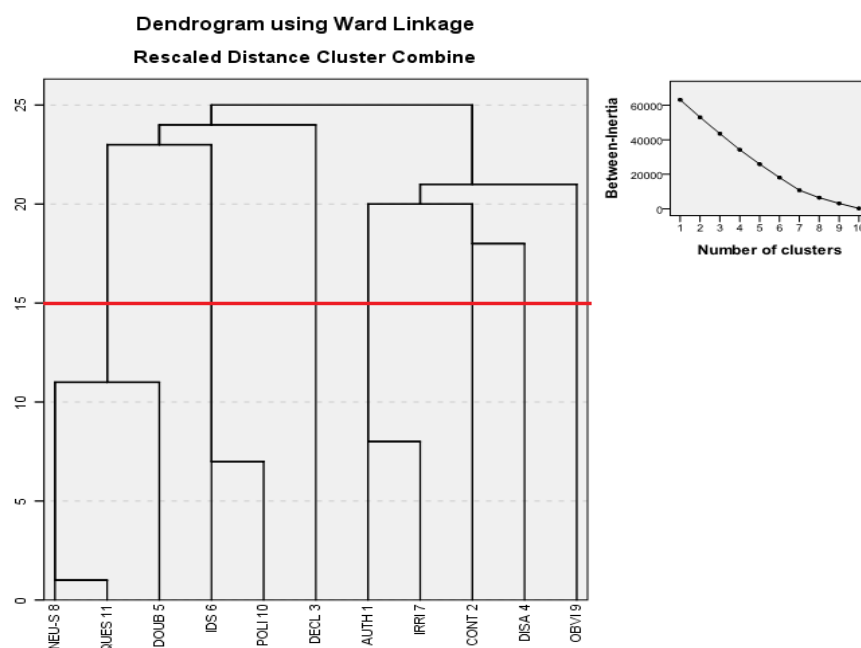


Figure 69 : Classification hiérarchique des 11 affects sociaux français pour les 14 sujets français dans la modalité « audio-vidéo ».

#### 4.4.5. Conclusion

Lors de cette étude de la perception multimodale des onze affects sociaux français, 14 sujets natifs ont participé à l'évaluation de la pertinence du corpus. Cette étude servira de référence à l'étude de la perception multimodale des mêmes affects sociaux auprès d'apprenants chinois du français, que nous présenterons dans la section suivante.

Un même énoncé a été exprimé avec les onze affects sociaux dans trois modalités (« audio seul », « vidéo seul » et « audio-vidéo »), puis ces 33 stimuli ont été jugés par des sujets natifs.

Dans un premier temps, nous avons vérifié l'importance des facteurs sur la réponse des sujets grâce à une analyse de régression logistique mixte. Selon le résultat, sont significatifs les facteurs d'affect social, de modalité et d'ordre de présentation des modalités, ainsi que leurs interactions d'ordre 2.

Ensuite, pour chaque affect social, des analyses de comparaison multiple ont été appliquées pour mettre en évidence les différences significatives des proportions de bonne réponse entre chacune des trois modalités, ainsi qu'entre les deux ordres de présentation des modalités. Quelques différences significatives sont observées pour le *mépris*, le *doute*, le *maternel* et l'*évidence*. Pour le *mépris*, le taux d'identification en

« audio seul » est significativement inférieur à ceux obtenus en « vidéo seul » et en « audio-vidéo » ( $p < 0,05$ ), ce qui montre l'importance de l'information visuelle dans la reconnaissance de l'expression ; pour le *doute*, le taux d'identification dans la condition audiovisuelle est significativement supérieur à celui de la condition audio seule ( $p < 0,05$ ) ; pour l'*évidence*, la différence entre la modalité audiovisuelle et la modalité visuelle seule est significative ( $p < 0,05$ ) ; pour le *maternel*, les taux d'identification dans les conditions audio seule et audiovisuelle sont significativement supérieurs à celui de la condition visuelle seule ( $p < 0,01$ ). Les indices auditifs contribuent ainsi de façon majeure à la reconnaissance de cette expression socio-affective.

En ce qui concerne l'effet de l'ordre de présentation des modalités, la majorité des affects sociaux ne montre pas de différences significatives entre les deux ordres de présentation des modalités sauf le *mépris*, le *doute* et l'*irritation* ( $p < 0,05$ ).

Au vu des matrices de confusion des affects sociaux dans chaque condition expérimentale, nous avons observé d'une part les principales confusions perceptives entre affects sociaux ainsi que l'orientation de ces confusions, et d'autre part les tendances du regroupement de ces affects sociaux selon leur proximité. L'*irritation* se rapproche de l'*autorité* dans les trois modalités, ce qui révèle de l'analogie entre ces deux affects sociaux en dépit de la modalité de présentation. La *question* n'est pas reconnue dans la condition visuelle seule (0%), et est principalement confondue avec la *déclaration* (57,1%). La *déception* est bien identifiée dans les trois modalités et sans confusions particulières avec les autres affects sociaux ; la *déclaration* est reconnue par l'ensemble de sujets en « audio-vidéo ».

Selon les résultats de la classification hiérarchique sur les mêmes matrices de confusion, et pour chaque modalité, les onze affects sociaux sont regroupés dans des catégories différentes en fonction des modalités considérées.

## **4.5. Perception multimodale des affects sociaux français par des apprenants chinois**

### **4.5.1. Introduction**

Après avoir validé le corpus des affects sociaux français multimodaux auprès de sujets natifs, nous procédons ensuite au comportement perceptif des apprenants chinois pour

les mêmes expressions socio-affectives présentées en multimodalité. L'objectif principal de cette étude consiste à vérifier le rôle et la contribution relative de l'audition et de la vision pour la perception des affects sociaux. Ceci est important pour l'apprentissage des langues étrangères, car les apprenants ont toujours besoin de se situer dans une situation face-à-face, dynamique et interactive. Parallèlement, les effets de l'ordre de la présentation des modalités et du niveau de français des apprenants chinois sont aussi examinés et les résultats de l'analyse sont comparés avec ceux des sujets français.

#### 4.5.2. Sujets

Deux groupes de 28 apprenants chinois de français de niveau débutant (niveau A) et intermédiaire (niveau B) ont été sélectionnés pour le test perceptif. Au sein de chaque groupe, les sujets ont été aléatoirement divisés en deux sous-groupes d'un nombre similaire. Les deux sous-groupes ont été distingués par l'ordre de présentation des modalités de stimulus. Le premier sous-groupe a passé les trois modalités dans l'ordre : « audio seul », « vidéo seul », puis « audio-vidéo », et le second groupe les a passées dans l'ordre : « vidéo seul », « audio seul », puis « audio-vidéo ».

Ainsi, 14 apprenants de niveau A (8 hommes et 6 femmes), d'âge moyen de 24.6 ans, ont été divisés en deux sous-groupes, chacun étant composé de 7 sujets, de même que les 14 apprenants de niveau B (4 hommes et 10 femmes), d'âge moyen de 25,1 ans.

Tous les participants font leurs études à Grenoble. Ils ont appris le français en Chine pendant au moins 500 heures avant de venir en France. Certains d'entre eux sont étudiants au Centre Universitaire d'Études Françaises (CUEF) de Grenoble, les autres sont étudiants en master à l'Université de Grenoble et à l'Institut national polytechnique de Grenoble. Ils ont passé un test de positionnement en français selon le CECRL (*Cadre Européen Commun de Référence pour les Langues*) au CUEF au début du nouveau semestre (pour les étudiants en licence) ou à l'entrée en master (pour les étudiants en master). Aucun d'entre eux n'a signalé de trouble auditif ou visuel.

#### 4.5.3. Sous-corpus

Le corpus audio-visuel des affects sociaux français validé par des sujets natifs a été retenu pour ce test perceptif.



#### 4.5.4. Protocole expérimental

Le protocole expérimental utilisé pour le test perceptif est le même que celui utilisé pour la perception multimodale des affects sociaux français par les 14 sujets natifs, présenté dans la partie 4.4.3.

#### 4.5.5. Analyse des résultats

##### 4.5.5.1. Facteurs influant la reconnaissance des affects sociaux

Une régression logistique mixte (voir supra) a été effectuée sur ces données pour étudier l'effet de plusieurs facteurs sur les réponses des sujets. Les facteurs fixés sont l'affect social présenté (11 affects sociaux), la modalité de l'expression socio-affective (« audio seul », « vidéo seul » et « audio-vidéo »), l'ordre de présentation des modalités (« audio seul » - « vidéo seul » - « audio-vidéo » ou « vidéo seul » - « audio seul » - « audio-vidéo ») et le niveau de français des apprenants chinois (niveau A et niveau B).

Une procédure descendante (*backward* en anglais) a été utilisée pour sélectionner les variables significatives. Le modèle initial de la procédure comprend les quatre facteurs et leurs interactions d'ordre 2. A chaque étape, la variable dont la valeur de  $p$  de l'effet associée au test de rapport de vraisemblance est la plus grande et supérieure à 0,05 est retirée du modèle. Le processus s'arrête lorsque toutes les variables sont retirées du modèle ou lorsque la valeur de  $p$  est inférieure à 0,05. Lorsqu'un facteur est impliqué dans une interaction significative, il ne peut pas être retiré du modèle même s'il est testé comme non significatif.

Le résultat de la sélection descendante est présenté par la table 21. La première colonne présente le nom de l'effet qui est retiré à chaque étape. Les trois colonnes suivantes donnent respectivement la valeur de la statistique du test, le nombre de degrés de liberté et la probabilité du test de rapport de vraisemblance. Selon le tableau, la valeur de  $p$  des effets présentés est supérieure à la probabilité, ils sont donc tous retirés du modèle. Autrement dit, les effets restants sont tous significatifs pour la variable réponse.

	LRT $\chi^2$	ddl	p
Modalité : Niveau de français	0,2897	2	0,8651
Modalité : Ordre de présentation des modalités	0,7035	2	0,7034
Ordre de présentation des modalités : Affect social	12,302	10	0,2654
Affect social : Niveau de français	15,03	10	0,131

Table 21 : Résultat de l'analyse de régression logistique mixte pas à pas descendante : les effets présentés dans le tableau sont testés comme **non significatifs** selon un test de rapport de vraisemblance ( $p > 0,05$ ).

Ainsi, les facteurs significatifs sont l'affect social, la modalité, l'ordre de présentation des modalités, le niveau de français des apprenants, ainsi que les interactions entre l'affect social et la modalité, et entre l'ordre de présentation des modalités et le niveau de français des apprenants.

#### 4.5.5.2. Effet de la modalité et de l'ordre de présentation des modalités

Ici, nous cherchons à mettre en avant, pour chaque affect social, les différences de proportions de bonne réponse entre les trois modalités et les deux ordres de présentation des modalités. Pour ce faire, nous avons appliqué la méthode de comparaison multiple, présentée en détail dans Hothorn et al. (2008). L'analyse a été réalisée à l'aide de la fonction « lsmeans » du package « lsmeans » du logiciel R.

- Effet de la modalité

Selon le résultat de la comparaison multiple, le *mépris*, la *déclaration*, le *doute*, le *maternel* et l'*évidence* montrent de différences significatives en modalités de présentation dans la reconnaissance. Pour le *mépris* et l'*évidence*, les taux de reconnaissance en « vidéo seul » et en « audio-vidéo » sont significativement supérieurs à ceux obtenus en condition « audio seul » ( $p < 0,01$ ) ; pour la *déclaration*, la différence significative existe entre la modalité audiovisuelle et visuelle seule ( $p < 0,05$ ) ; pour le *doute*, la progression du taux de reconnaissance est significative entre « audio seul » et « audio-vidéo », ainsi qu'entre « vidéo seul » et « audio-vidéo » ( $p < 0,05$ ) ; pour le *maternel*, le taux d'identification en « vidéo seul » est significativement inférieur à ceux observés en « audio seul » et en « audio-vidéo » ( $p < 0,01$ ) .

Ce résultat est illustré par la figure 70, qui représente le taux de reconnaissance moyen obtenu pour chacun des affects sociaux dans chacune des trois modalités. Presque tous les affects sociaux sont reconnus par les apprenants chinois dans toutes les modalités (le

taux d'identification est supérieur au seuil du hasard), à l'exception de la *question* (QUES) et de l'exclamation de *surprise neutre* (NEU-S) (0%) en vidéo seul. Les résultats obtenus en modalité audiovisuelle montrent des meilleurs taux d'identification pour tous les affects sociaux à l'exception de la *question*. Les expressions de *mépris* (CONT) et d'*évidence* (OBVI) sont plus dépendantes de l'information visuelle que de l'information auditive ; le *maternel* (IDS) a plus recours à l'information auditive, ainsi que la *question*. Pour le reste des affects sociaux : l'*autorité* (AUTH), la *déclaration* (DECL), la *déception* (DISA), l'*irritation* (IRRI), la *politesse* (POLI) et le *doute* (DOUB), l'information auditive et visuelle coopèrent et se complètent l'une avec l'autre dans l'expression. Parmi ces affects sociaux, le *doute* (DOUB) montre le plus de synergie audiovisuelle.

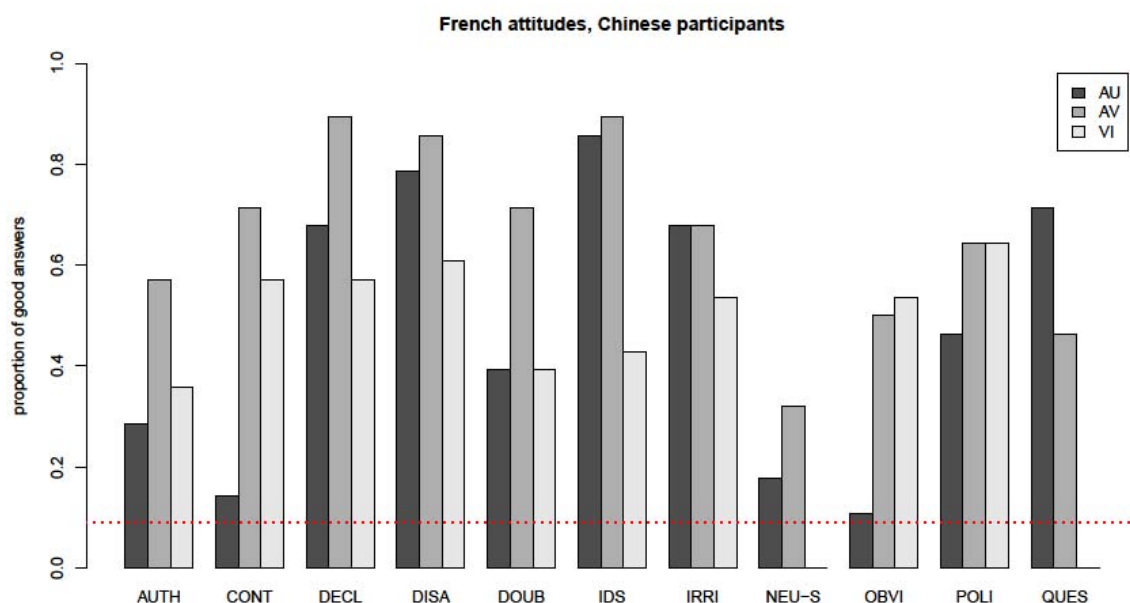


Figure 70 : Effet de la modalité sur les taux d'identification des 11 affects sociaux français pour les 28 apprenants chinois de français. Les affects sociaux sont présentés en abscisse et les scores de reconnaissance en ordonnée. La ligne rouge pointillée indique le seuil du hasard (9%).

- Effet de l'ordre de présentation des modalités

Selon le résultat de la comparaison multiple, l'interaction entre l'ordre de présentation des modalités et le niveau de français des apprenants a un effet significatif sur la reconnaissance des expressions socio-affectives : les apprenants chinois de niveau B identifient significativement mieux les affects sociaux que les apprenants de niveau A dans l'ordre de « vidéo seul » - « audio seul » - « audio-vidéo » ( $p < 0,05$ ). Ce résultat est représenté par la figure 71.

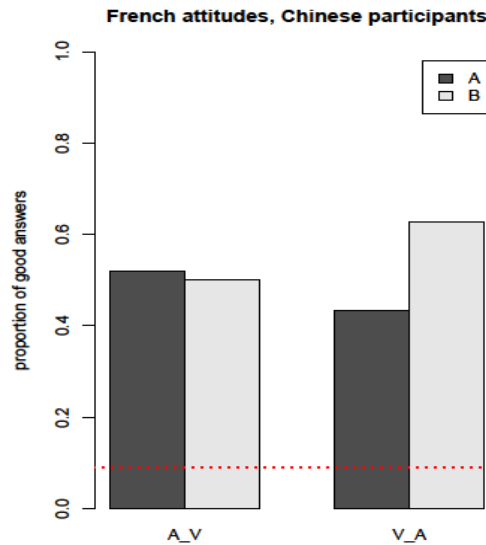


Figure 71 : Effet de l'interaction entre l'ordre de présentation des modalités et le niveau de français sur les taux d'identification des 11 affects sociaux français pour les 28 apprenants chinois. Les affects sociaux sont présentés en abscisse et les scores de reconnaissance en ordonnée. La ligne rouge pointillée indique le seuil du hasard (9%).

#### 4.5.5.3. Confusion perçue des affects sociaux

L'analyse des confusions entre affects sociaux a été réalisée pour les 28 apprenants chinois de français, sur la base de la matrice de confusions de chaque modalité. Les résultats des deux groupes d'apprenants ont été mélangés. Cette analyse a été effectuée en deux étapes : la première étape consiste à discuter les confusions les plus importantes entre affects sociaux et l'orientation de ces confusions, ce qui correspond bien à la problématique de la perception de ces affects sociaux dans une modalité donnée ; la seconde étape est basée sur une analyse de classification hiérarchique de ces expressions socio-affectives dont le but est d'extraire les tendances générales du regroupement des affects sociaux en considération de l'ensemble de confusions perceptives apparues lors de l'expérience.

Lors de l'analyse de classification hiérarchique, le carré de la distance Euclidienne a été utilisé comme distance entre les observations, qui ont été alors regroupées en classes différentes selon le critère de variance minimale de Wald. En vue de déterminer le nombre optimal de groupes d'affects sociaux sur l'arbre hiérarchique, nous nous sommes référés au critère de réduction de la variance (*elbow criterion* en anglais)

expliquée par chaque groupe. Selon ces résultats, les onze affects sociaux peuvent être catégorisés en groupes qui varient selon la modalité considérée.

- Modalité « audio seul »

La figure 72 représente les principales confusions perceptives entre les affects sociaux pour les 28 apprenants chinois de français dans la condition audio seule, obtenues à partir de la matrice de confusion. Les confusions deux fois supérieures au hasard (9%) sont proposées.

Il est à noter que la *politesse* (POLI) est confondue avec la *déclaration* (DECL) et le *maternel* (IDS) ; le *mépris* (CONT) est parfois mélangé avec la *déclaration* ; l'*autorité* (AUTH) est confondue avec l'*irritation* (IRRI) avec un pourcentage important (60,8%) ; l'*évidence* (OBVI) se reporte sur l'exclamation de *surprise neutre* (NEU-S) et sur le *doute* (DOUB), tandis que le dernier montre des confusions vers la *question* (QUES) et vers l'exclamation de *surprise neutre* (NEU-S). La *déception* (DISA) est bien reconnue et ne montre pas de confusions particulières avec les autres affects sociaux.

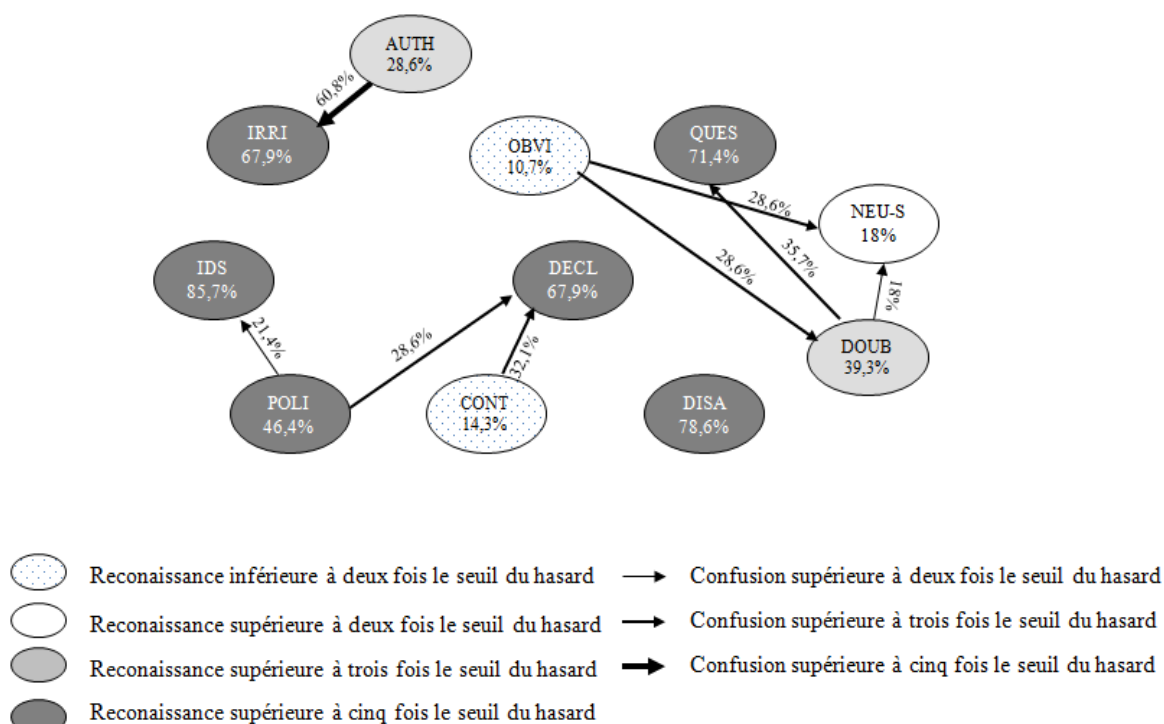


Figure 72 : Confusions des 11 affects sociaux français pour les 28 apprenants chinois de français dans la modalité « audio seul ». Les pourcentages d'identification sont indiqués dans les ovales et les taux de confusions sur les flèches. Seules les confusions qui affichent des scores supérieurs à deux fois le seuil du hasard sont rapportées.

Par la suite, nous avons mené une analyse de classification hiérarchique sur la base de la même matrice de confusions, afin d'identifier des catégories de ces affects sociaux pour les apprenants chinois dans la condition « audio seul ». Le dendrogramme de la figure 73 représente les résultats de cette analyse. Le graphique situé en haut à droite du dendrogramme illustre la perte d'inertie en fonction du nombre de groupes, à partir duquel on peut conclure qu'il est raisonnable de considérer cinq catégories d'affects sociaux.

Le *maternel* (IDS) et la *déception* (DISA) sont bien reconnus par les sujets et forment des isolats dans cette condition expérimentale. Les autres affects sociaux sont répartis en trois catégories distinctes : la *déclaration* (DECL), la *politesse* (POLI) et le *mépris* (CONT) sont regroupés dans la première catégorie ; l'*autorité* (AUTH) et l'*irritation* (IRRI), deux affects sociaux d'imposition sont réunies dans la deuxième catégorie ; la troisième catégorie contient le *doute* (DOUB), l'exclamation de *surprise neutre* (NEU-S), l'*évidence* (OBVI) et la *question* (QUES).

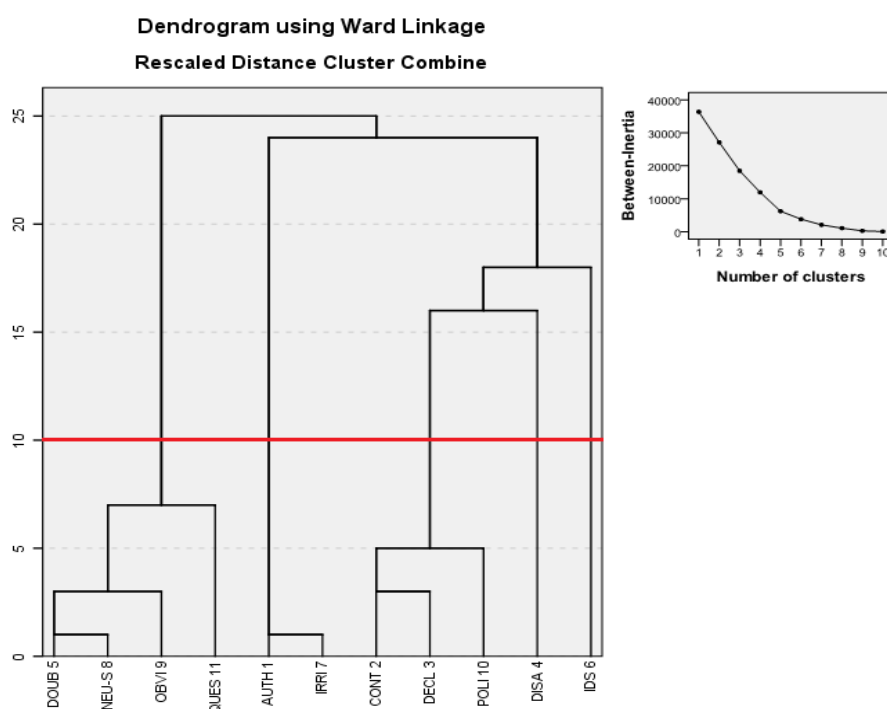


Figure 73 : Classification hiérarchique des 11 affects sociaux français pour les 28 apprenants chinois du français dans la modalité « audio seul ».

- Modalité « vidéo seul »

Les principales confusions entre les affects sociaux français pour tous les apprenants chinois dans la modalité « vidéo seul » sont schématisées par la figure 74. Beaucoup de confusions sont observées dans cette condition expérimentale. Le *mépris* (CONT) et l'*évidence* (OBVI) sont confondus avec l'*irritation* (IRRI); l'*irritation* et l'*autorité* (AUTH) sont confondus réciproquement, ainsi que le *maternel* (IDS) et la *politesse* (POLI); l'*autorité* se reporte aussi sur l'*évidence*; le *doute* est mélangé avec la *déception* (DISA) avec la *question* (QUES); l'exclamation de *surprise neutre* (NEU-S) et la *question* ne sont pas reconnues par les apprenants chinois dans la modalité visuelle seule (0%), et toutes les deux attitudes sont largement confondues avec la *déclaration* (50% pour l'exclamation de *surprise neutre*, 53,6% pour la *question*).

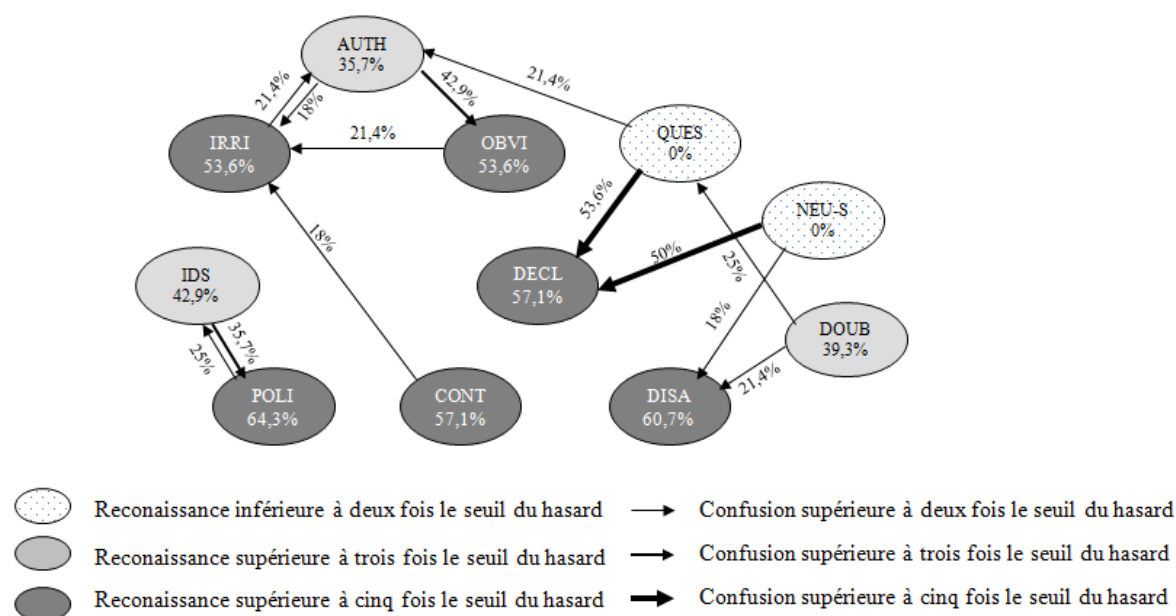


Figure 74 : Confusions des 11 affects sociaux français pour les 28 apprenants chinois de français dans la modalité « vidéo seul ». Les pourcentages d'identification sont indiqués dans les ovales et les taux de confusions sur les flèches. Seuls les taux supérieurs à deux fois le seuil du hasard sont rapportés.

La figure 75 illustre les résultats de la classification hiérarchique sur les mêmes données dans la condition « vidéo seul ». Les onze affects sociaux sont regroupés en quatre catégories sur la base de leur proximité dans cette condition expérimentale.

La première catégorie regroupe les expressions de *déclaration* (DECL), de *question* (QUES) et d'exclamation de *surprise neutre* (NEU-S), qui sont marquées par une valence neutre du locuteur à l'égard de ce qu'il vient de dire. Comme les sujets français, les

apprenants chinois jugent similaires l'*autorité* (AUTH), l'*évidence* (OBVI) et l'*irritation* (IRRI) et les regroupent dans la deuxième catégorie. La troisième catégorie contient le *mépris* (CONT), la *déception* (DISA) et le *doute* (DOUB), et la dernière est composée du *maternel* (IDS) et de la *politesse* (POLI).

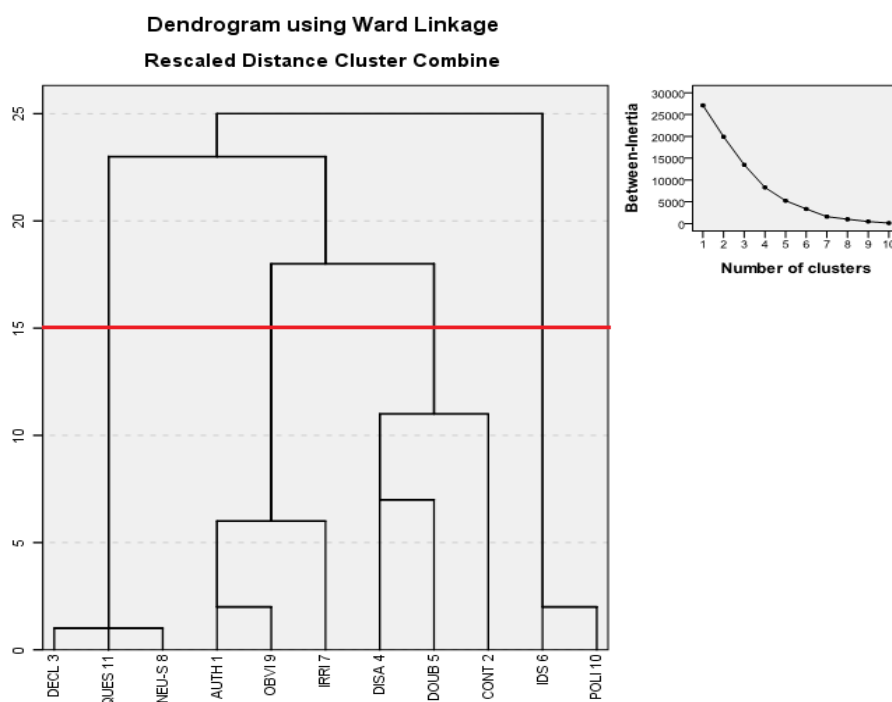


Figure 75 : Classification hiérarchique des 11 affects sociaux français pour les 28 apprenants chinois de français dans la modalité « vidéo seul ».

- Modalité « audio-vidéo »

La figure 76 représente les taux d'identification et les confusions des affects sociaux dans la modalité audiovisuelle. Les sujets reconnaissent mieux les affects sociaux dans cette condition expérimentale grâce aux synergies entre les deux modalités. Presque tous les affects sociaux sont reconnus cinq fois au-dessus du hasard, à l'exception de l'exclamation de *surprise neutre* (NEU-S), qui est cependant correctement reconnue (32,1%). La *déclaration* (DECL) et le *maternel* (IDS) reçoivent le meilleur score de reconnaissance (89,3%). L'*autorité* (AUTH) est confondue avec l'*évidence* (OBVI) et l'*irritation* (IRRI) ; le *doute* (DOUB) et l'exclamation de *surprise neutre* (NEU-S) se reportent sur la *question* (QUES), elle-même étant confondue principalement avec la *déclaration* (DECL) ; l'expression de *mépris* (CONT) est mélangée avec celle de *déception* (DISA), ainsi que la *politesse* (POLI) avec le *maternel* (IDS).



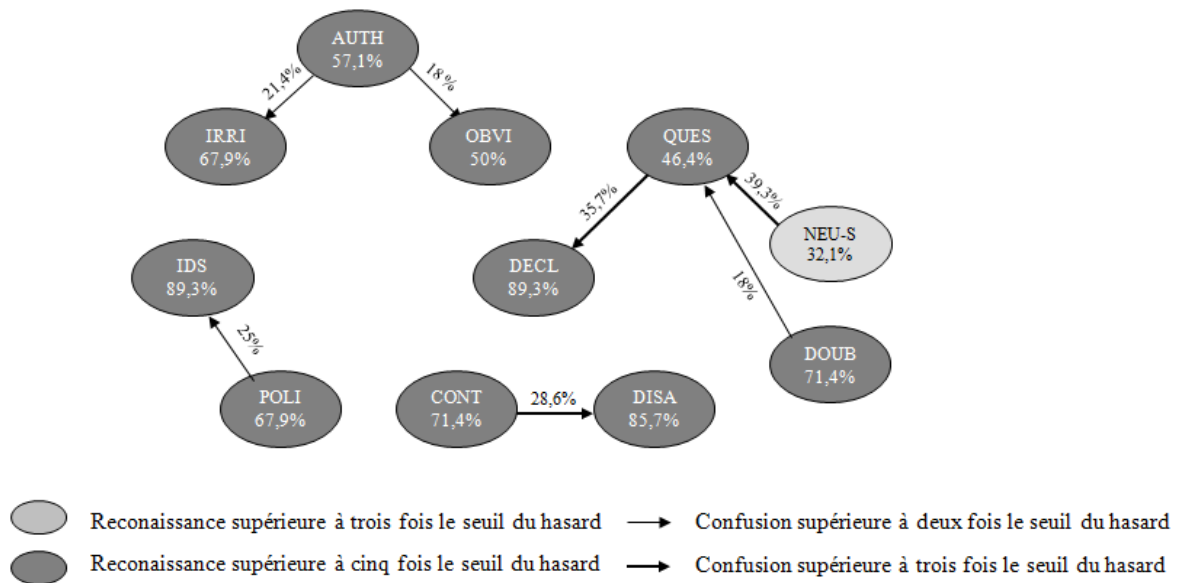


Figure 76 : Confusions des 11 affects sociaux français pour les 28 apprenants chinois de français dans la modalité « audio-vidéo ». Les pourcentages d'identification sont indiqués dans les ovales et les taux de confusions sur les flèches. Seules les confusions qui affichent des scores supérieurs à deux fois le seuil du hasard sont rapportées.

Les résultats de la catégorisation de ces affects sociaux dans la condition « audio-vidéo » sont illustrés par la figure 77. Les onze affects sociaux se répartissent en cinq catégories dans cette condition expérimentale.

Selon les résultats, la première catégorie regroupe le *doute* (DOUB), la *question* (QUES) et l'exclamation de *surprise neutre* (NEU-S) ; la deuxième catégorie contient l'*autorité* (AUTH), l'*évidence* (OBVI) et l'*irritation* (IRRI) ; le *mépris* (CONT) et la *déception* (DISA) sont classés dans la troisième catégorie ; la quatrième catégorie est composée de la *politesse* (POLI) et du *maternel* (IDS). La *déclaration* (DECL) est bien reconnue et ne montre pas de confusion particulière avec les autres affects sociaux dans cette condition expérimentale.

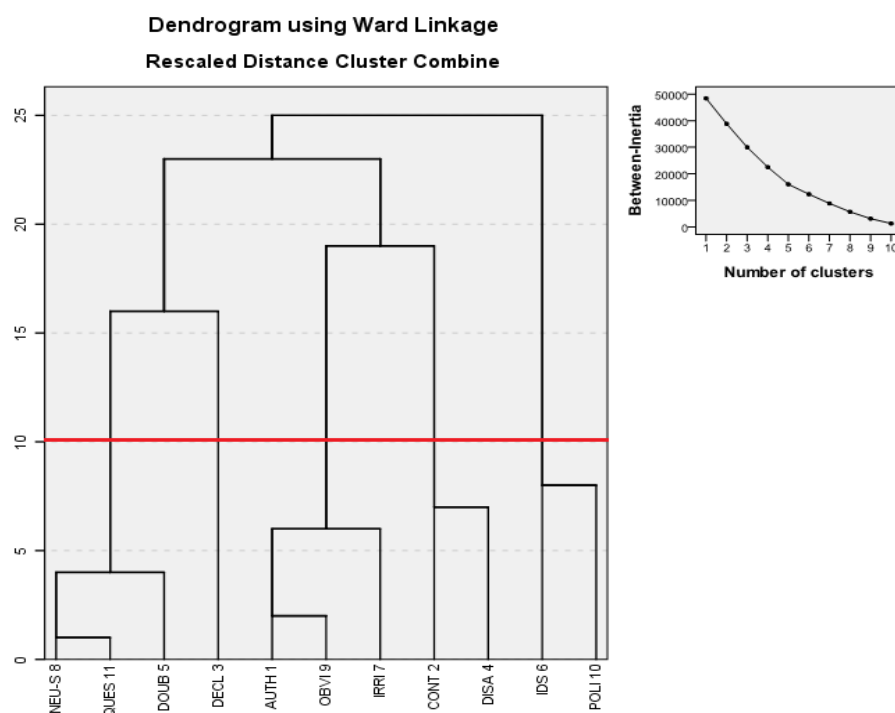
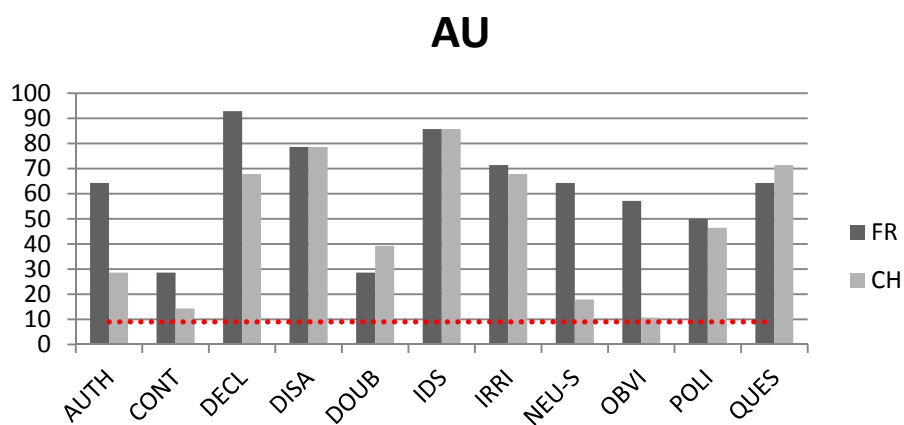


Figure 77 : Classification hiérarchique des 11 affects sociaux français pour les 28 apprenants chinois du français dans la modalité « audio vidéo ».

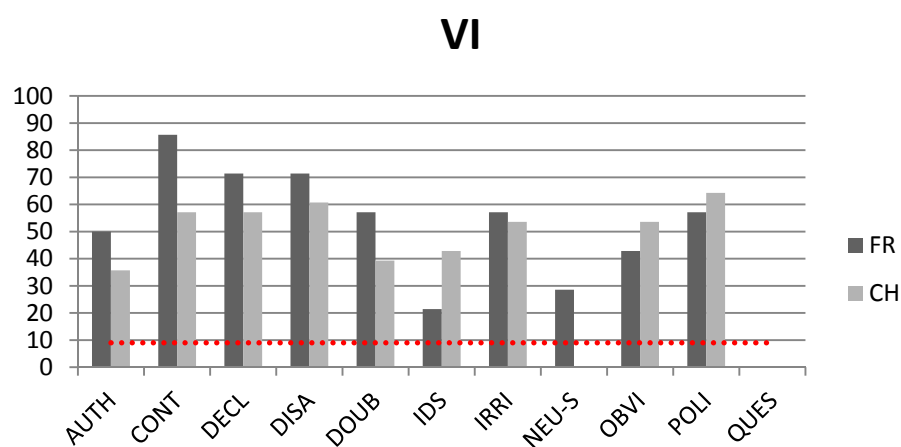
#### 4.5.5.4. Comparaison entre les apprenants chinois et les sujets natifs

Afin d'éclairer les essentielles différences entre les sujets natifs et les apprenants chinois en matière de perception des socio-affects français, nous comparons ici les résultats de reconnaissance des affects sociaux français obtenus auprès d'apprenants chinois avec ceux des sujets français natifs.

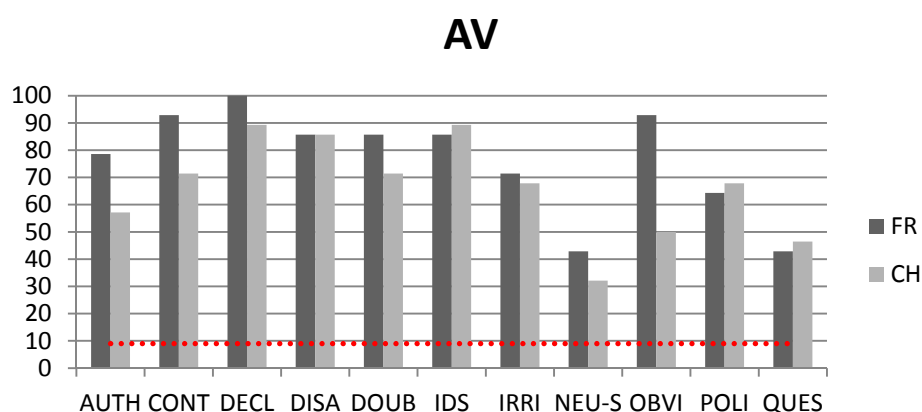
Au niveau du taux d'identification des affects sociaux dans les trois modalités, illustré par la figure 78, il est constaté que, dans la condition audio seule (figure 78, a), les sujets natifs identifient mieux l'*autorité* (AUTH), l'exclamation de *surprise neutre* (NEU-S) et l'*évidence* (OBVI) que les apprenants chinois. Dans la condition visuelle seule (figure 78, b), les apprenants ne reconnaissent pas l'expression de l'exclamation de *surprise neutre* ni la *question* (QUES) (0%). Il est à noter que la *question* n'est pas reconnue non plus par les sujets natifs. Les apprenants chinois obtiennent un meilleur score pour le *maternel* (IDS) que les sujets natifs en vidéo seul. Dans la modalité audiovisuelle (figure 78, c), les deux groupes de sujets perçoivent les affects sociaux de manière similaire, à l'exception de l'*évidence*, qui montre un décalage perceptif évident entre les sujets natifs et non natifs.



(a) Modalité audio seule



(b) Modalité visuelle seule



(c) Modalité audio-visuelle

Figure 78 : Taux de reconnaissance de 11 affects sociaux français pour les sujets natifs et les apprenants chinois dans les trois modalités : « audio seul » (haut) ; « vidéo seul » (milieu) ; « audio-vidéo » (bas).

En ce qui concerne l'effet de la modalité sur les expressions de onze affects sociaux français pour les sujets natifs et les apprenants chinois du français, globalement, les expressions sont mieux reconnues par tous les sujets quand les informations auditives et visuelles sont mêlées. Cependant, l'audition et la vision ne jouent pas un rôle identique dans l'expression de certains socio-affects : pour le *mépris*, tous les sujets ont plus recours à l'indice visuel qu'à l'indice auditif ; pour le *maternel*, le message affectif est plus véhiculé par l'information auditive que par l'information visuelle ; enfin, le *doute* est plus dépendant de l'information auditive que de l'information visuelle pour les apprenants chinois, mais est dépendant de la combinaison des deux modalités par les sujets natifs.

Quant à l'effet de l'ordre de présentation des modalités, selon le résultat de l'analyse statistique, ce facteur a un effet significatif sur la reconnaissance des expressions de *mépris*, de *doute* et d'*irritation* pour les sujets natifs. Pour les apprenants chinois, son interaction avec le niveau des apprenants montre un effet significatif sur la reconnaissance des affects sociaux. Lorsque les stimuli sont présentés dans l'ordre de « vidéo seul » - « audio seul » - « audio-vidéo », les apprenants de niveau intermédiaire identifient significativement mieux les expressions socio-affectives que les apprenants de niveau débutant.

Enfin, en matière de confusions perceptives entre affects sociaux, beaucoup de points communs entre les sujets français et chinois sont à noter en fonction de la modalité. En « audio seul », l'*autorité* est confondue avec l'*irritation* ; la *politesse* et le *mépris* se reportent sur la *déclaration* ; le *doute* est parfois mélangé avec la *question* et l'exclamation de *surprise neutre*. En « vidéo seul », l'*irritation* est considérée similaire à l'*autorité* ; le *maternel* est confondu avec la *politesse* ; l'*autorité* est mélangée avec l'*évidence* ; la *question*, attitude non reconnue ni par les sujets natifs ni par les apprenants chinois (0%), est largement confondue avec la *déclaration* dans cette modalité. En « audio-vidéo », pour tous les sujets, la *politesse* est confondue avec le *maternel*, l'*irritation* avec l'*autorité*, ainsi que l'exclamation de *surprise neutre* avec la *question*.

#### 4.5.6. Conclusion

Dans cette étude de la perception multimodale de onze affects sociaux français, un seul énoncé a été exprimé avec les affects sociaux dans trois modalités (« audio seul », « vidéo seul » et « audio-vidéo »). Ces 33 stimuli ont été jugés par 28 apprenants chinois de français de niveau débutant (niveau A) et intermédiaire (niveau B).

Dans un premier temps, nous avons vérifié l'importance des facteurs sur la réponse des sujets grâce à une analyse de régression logistique mixte. Selon le résultat, les facteurs significatifs sont l'affect social, la modalité, l'ordre de présentation des modalités, le niveau de français des apprenants, ainsi que l'interaction entre affect social et modalité, et celle entre l'ordre de présentation des modalités et le niveau de français des apprenants chinois.

Ensuite, pour chaque affect social, des analyses de comparaison multiple ont été appliquées pour mettre en évidence les différences significatives de proportions de bonne réponse entre les trois modalités, ainsi qu'entre les deux ordres de présentation des modalités pour chaque groupe de sujets. Selon les résultats, le *mépris* et l'*évidence* sont significativement mieux reconnus en « vidéo seul » et « audio-vidéo » qu'en « audio seul » ( $p < 0,01$ ), elles sont donc plus dépendants de l'information visuelle que de l'information auditive; le *maternel*, dont le taux d'identification en « vidéo seul » est significativement inférieur à ceux observés en « audio seul » et en « audio-vidéo » ( $p < 0,01$ ), est plus dépendant de l'information auditive ; le *doute* est significativement mieux reconnu en « audio-vidéo » qu'en « audio seul » et « vidéo seul » ( $p < 0,05$ ), ce qui démontre l'importance de la synergie entre ces deux modalités dans l'expression.

Parallèlement, l'interaction entre l'ordre de présentation des modalités et le niveau de français des apprenants chinois a également un effet significatif sur la reconnaissance des affects sociaux. Cependant, il ne s'agit que de l'ordre de « vidéo seul » - « audio seul » - « audio-vidéo », dans lequel les apprenants de niveau B identifient significativement mieux les expressions socio-affectives que les apprenants de niveau A.

Par la suite, une analyse de confusions perceptives des affects sociaux dans chaque condition expérimentale, basée sur les matrices de confusion, a été réalisée pour mettre en évidence les comportements perceptifs des apprenants chinois dans chacune des modalités. Cette analyse a été effectuée en deux étapes : la première étape consiste à

examiner les confusions principales et leur orientation ; la seconde étape a pour but d'extraire les tendances générales du regroupement des affects sociaux en considération de l'ensemble de confusions perceptives apparues lors de l'expérience. Selon les résultats, l'*autorité* se reporte sur l'*irritation* dans toutes les modalités, ainsi que la *politesse* sur le *maternel* et le *doute* sur la *question* ; la *question* est confondue avec la *déclaration* dans la condition visuelle seule et audiovisuelle, mais elle ne montre pas de confusions particulières avec les autres affects sociaux dans la condition audio seule. Cela indique que la prosodie joue un rôle proéminent dans l'expression interrogative en français. Au point de vue de la catégorisation de ces affects sociaux, le regroupement de l'*autorité* et l'*irritation* est observé dans les trois modalités ; le *doute*, la *question* et l'exclamation de *surprise neutre* sont regroupés en « audio seul » et « audio-vidéo », mais ils sont répartis dans différents groupes en « vidéo seul » : la *question* et l'exclamation de *surprise neutre* sont regroupées avec la *déclaration* et le *doute* avec la *déception* et le *mépris*.

Enfin, en comparant le comportement perceptif des apprenants chinois avec celui des sujets français pour les mêmes stimuli, nous constatons que les sujets natifs et les apprenants chinois confondent certains affects sociaux de manière similaire dans chacune des modalités. D'un autre côté, la modalité a un impact similaire sur la reconnaissance de certains affects sociaux pour les deux groupes de sujets : le *mépris* dépend plus de l'information visuelle que de l'information auditive ; le *maternel* est plus dépendant de l'information auditive ; le *doute* s'appuie plutôt sur la combinaison des deux modalités. Par contre, l'*évidence* montre des différences entre les sujets natifs et les apprenants chinois, car pour l'identifier, les apprenants chinois ont plus recours à l'information visuelle qu'à l'information auditive, alors que les sujets natifs ne montrent pas de préférence lors de leur perception.

## **4.6. Perception multimodale des affects sociaux chinois par des apprenants français**

### **4.6.1. Introduction**

Dans le chapitre 2, nous avons examiné la perception multimodale de onze affects sociaux chinois auprès de sujets chinois natifs. Cette expérience nous sert d'étude de référence pour cette étude de la perception des expressions socio-affectives chinoises

auprès d'apprenants français. Dans cette étude, nous examinons le comportement perceptif des apprenants français pour les mêmes affects sociaux chinois présentés en modalités différentes. Les effets de la modalité, de l'ordre de présentation des modalités et du niveau de chinois des apprenants français sont mis en considération. Les résultats sont par la suite comparés avec ceux obtenus pour les sujets natifs chinois (voir partie 2.5).

#### 4.6.2. Sous-corpus

Les mêmes stimuli utilisés lors de l'expérience des affects sociaux chinois pour les sujets natifs ont été retenus pour ce test perceptif proposé aux sujets français: un énoncé de quatre syllabes réalisé avec les onze affects sociaux selon les trois modalités « audio seul », « vidéo seul » et « audio-vidéo ». L'énoncé sélectionné est « 四天三夜 », « si4 tian1 san1 ye4 » en pinyin et « quatre jours trois nuits » en français. Les affects sociaux choisis sont la *déclaration* (DECL), l'*évidence* (OBVI), le *doute* (DOUB), la *question* (QUES), l'*autorité* (AUTH), l'exclamation de *surprise neutre* (NEU-S), la *déception* (DISA), le *mépris* (CONT), l'*irritation* (IRRI), la *politesse* (POLI) et le *maternel* (IDS).

#### 4.6.3. Sujets

28 apprenants français de chinois mandarin de niveau débutant (niveau A) et intermédiaire (niveau B) ont participé à cette expérience : 14 apprenants de niveau A (3 hommes et 11 femmes), d'âge moyen de 21,5 ans, et 14 apprenants de niveau B (5 hommes et 9 femmes), d'âge moyen de 22,3 ans.

Tous les participants font leurs études à Grenoble et apprennent le chinois mandarin au département de LANSAD (Langues pour Spécialistes d'Autres Disciplines) de l'Université Stendhal – Grenoble III. Ils ont passé un test de positionnement en chinois mandarin selon le CECRL (*Cadre Européen Commun de Référence pour les Langues*) au début du semestre. Aucun d'entre eux n'a signalé de trouble auditif ou visuel.

Durant le test, les sujets de chaque niveau de chinois ont été aléatoirement divisés en deux groupes. Les modalités ont été présentées à chacun des groupes selon un ordre différent : le premier groupe (7 sujets) a passé les trois modalités dans l'ordre « audio seul », « vidéo seul », puis « audio-vidéo » ; le second groupe (7 sujets) a passé les trois modalités dans l'ordre « vidéo seul », « audio seul », puis « audio-vidéo ».

#### 4.6.4. Protocole expérimental

Le protocole expérimental utilisé pour le test perceptif est le même que celui utilisé pour la perception multimodale des affects sociaux chinois par des sujets chinois (voir partie 2.5), et des affects sociaux français par des sujets français et des apprenants chinois, présenté dans la partie 4.4.3.

#### 4.6.5. Analyse des résultats

##### 4.6.5.1. Facteurs influant la reconnaissance des affects sociaux

Une régression logistique a été effectuée sur les données recueillies lors de l'expérience pour examiner l'effet de plusieurs facteurs sur la variable réponse binaire. Les facteurs fixés sont l'affect social présenté (11 affects sociaux), la modalité de l'expression attitudinale (« audio seul », « vidéo seul » ou « audio-vidéo »), l'ordre de présentation des modalités (« audio seul » - « vidéo seul » - « audio-vidéo » ou « vidéo seul » - « audio seul » - « audio-vidéo »), et le niveau de chinois des apprenants français (niveau A, niveau B).

Une procédure descendante (*backward* en anglais) a été utilisée pour sélectionner les variables significatives. Le modèle initial de la procédure comprend les 4 facteurs et leurs interactions d'ordre 2. A chaque étape, la variable dont la valeur de  $p$  de l'effet associée au test de rapport de vraisemblance est la plus grande et supérieure à 0,05 est retirée du modèle. Le processus s'arrête lorsque toutes les variables sont retirées du modèle ou lorsque la valeur de  $p$  est inférieure à 0,05. Lorsqu'un facteur est impliqué dans une interaction significative, il ne peut pas être retiré du modèle même s'il est testé comme non significatif.

La table 22 donne le résultat de la sélection descendante. La première colonne présente le nom de l'effet qui est retiré à chaque étape. Les trois colonnes suivantes donnent respectivement la valeur de la statistique du test, le nombre de degrés de libertés et la probabilité du test de rapport de vraisemblance. Selon le tableau, la valeur de  $p$  des effets présentés est supérieure à la probabilité, ils sont donc tous retirés du modèle. Autrement dit, les effets restants ont un effet significatif sur la variable réponse.



	LRT $\chi^2$	ddl	<i>p</i>
Ordre de présentation des modalités : Attitude	5,8128	10	0,8307
Modalité : Niveau de chinois	0,5076	2	0,7758
Ordre de présentation des modalités : niveau de chinois	0,6679	1	0,4138
Attitude : Niveau de chinois	10,461	10	0,401
Modalité : Ordre de présentation des modalités	3,8693	2	0,1445
Ordre de présentation des modalités	0,6694	1	0,4133
Niveau de chinois	0,0242	1	0,8764

Table 22 : Résultat de l'analyse de régression logistique mixte pas à pas descendante : les effets présentés dans le tableau sont testés *non significatifs* au test de rapport de vraisemblance ( $p > 0,05$ ).

Ainsi, les facteurs significatifs sont l'affect social, la modalité, ainsi que l'interaction entre l'affect social et la modalité. L'ordre de présentation des modalités et le niveau de chinois des apprenants français n'ont pas d'effet significatif sur les réponses des sujets.

#### 4.6.5.2. Effet de la modalité

Afin de déterminer l'effet de la modalité sur les choix catégoriels moyens des sujets pour les affects sociaux, nous avons effectué une analyse de comparaison multiple en calculant, pour chaque affect social, les différences de proportions de bonne réponse entre les trois modalités. L'analyse a été réalisée à l'aide de la fonction « lsmeans » du package « lsmeans » du logiciel R.

Selon le résultat de cette analyse, la plupart des affects sociaux ne montrent pas de différences significatives entre les modalités de présentation en matière de reconnaissance, sauf la *déception*, le *doute*, le *maternel* et l'*évidence*. Pour la *déception*, le taux de reconnaissance en « audio-vidéo » est significativement plus élevé qu'en « vidéo seul » ( $p < 0,05$ ) ; pour le *doute*, la progression du taux de reconnaissance est très significative entre « audio seul » et « audio-vidéo » ( $p < 0,01$ ) ; pour le *maternel*, les taux de reconnaissance en modalité audiovisuelle et en modalité audio seule sont significativement supérieurs à celui observé en modalité visuelle seule ( $p < 0,01$  ;  $p < 0,05$ ) ; l'*évidence* montre un taux significativement plus élevé en modalité audiovisuelle qu'en modalité audio seule ( $p < 0,01$ ).

Ce résultat est illustré par la figure 79, qui représente le taux de reconnaissance moyen obtenu pour chacun des affects sociaux dans chacune des trois modalités. Généralement,

presque tous les affects sociaux chinois sont reconnus par les apprenants français dans toutes les modalités (le taux d'identification est supérieur au seuil du hasard), sauf la *question* (QUES) et l'exclamation de *surprise neutre* (NEU-S) en « vidéo seul », et la *politesse* (POLI) en « audio seul ». Les résultats obtenus en modalité audiovisuelle montrent des meilleurs scores d'identification pour tous les affects sociaux sauf pour la *question* (QUES). Par rapport à l'information auditive, l'information visuelle joue un rôle majeur dans la reconnaissance de l'expression de *doute* (DOUB) et d'*évidence* (OBVI). A l'inverse, pour l'expression de *maternel* (IDS), les sujets ont plus recours à l'information auditive qu'à l'information visuelle. L'expression de *politesse* n'est pas reconnue par les sujets français en « audio seul » (0%). Pour le reste des affects sociaux : l'*autorité* (AUTH), le *mépris* (CONT), la *déclaration* (DECL), la *déception* (DISA), la *question*, l'*irritation* (IRRI) et l'exclamation de *surprise neutre* (NEU-S), l'information auditive et visuelle coopèrent et se complètent l'une avec l'autre dans l'expression.

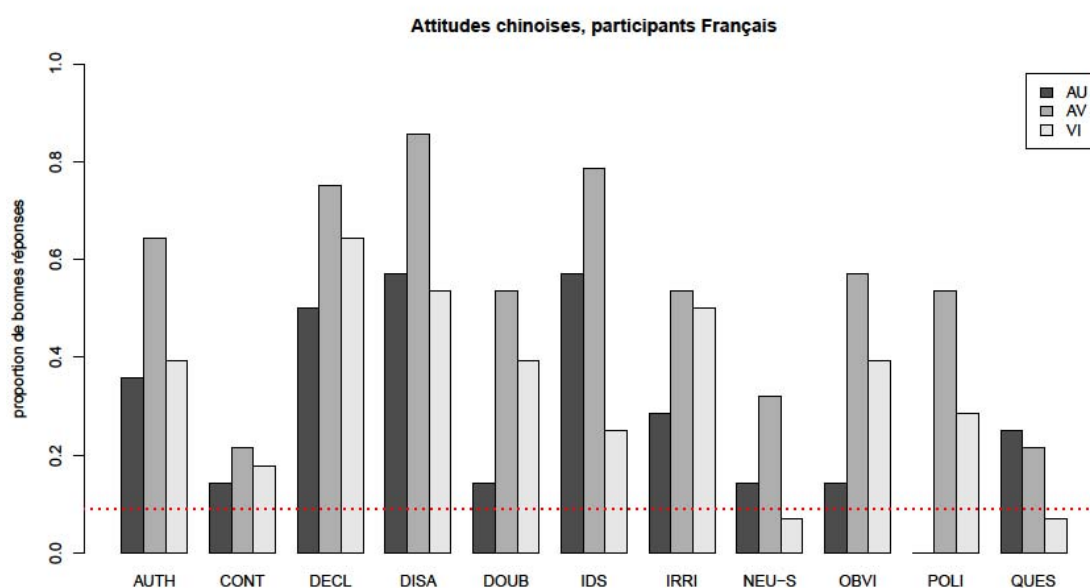


Figure 79 : Effet de la modalité sur les taux d'identification des 11 affects sociaux chinois pour les 28 apprenants français de chinois mandarin. Les affects sociaux sont présentés en abscisse et les scores de reconnaissance en ordonnée. La ligne rouge pointillée indique le seuil du hasard (9%).

#### 4.6.5.3. Confusion perçue des affects sociaux

L'analyse des confusions entre les affects sociaux est réalisée pour les 28 apprenants français du chinois mandarin, sur la base de la matrice de confusion de chaque modalité. Les résultats des deux groupes de sujets ont été mélangés, car le facteur de niveau de

chinois n'a pas induit d'effet significatif sur la reconnaissance des affects sociaux d'après le résultat de la régression logistique.

Lors de l'analyse de classification hiérarchique, le carré de la distance Euclidienne a été utilisé comme distance entre les observations, qui ont été alors regroupées en classes différentes selon le critère de variance minimale de Wald. En vue de déterminer le nombre optimal de groupes sur l'arbre hiérarchique, nous nous sommes référés au critère de réduction de la variance (« elbow criterion » en anglais) expliquée par chaque groupe (Soni Madhulatha, 2012). Selon les résultats, les onze affects sociaux peuvent être répartis en groupes qui varient selon la modalité considérée.

- Modalité « audio seul »

La figure 80 illustre les principales confusions perceptives entre affects sociaux pour les 28 apprenants français dans la condition audio seule, obtenues à partir de la matrice de confusion. Les confusions deux fois supérieures au hasard (9%) sont proposées.

Selon le schéma, l'*évidence* (OBVI), l'*autorité* (AUTH) et la *politesse* (POLI) sont principalement confondues avec la *déclaration* (DECL), qui montre quelques confusions avec l'*évidence* (OBVI). L'*autorité* montre des confusions réciproques avec l'*irritation* (IRRI). L'exclamation de *surprise neutre* (NEU-S), le *mépris* (CONT), le *doute* (DOUB) et la *question* (QUES) sont principalement confondus entre eux. Il est intéressant de remarquer que la *politesse* n'est pas reconnue dans cette modalité auditive (0%) et est largement confondue avec la *déclaration* (57,1%). Ce phénomène a été aussi observé dans les réponses des sujets natifs, qui ont principalement confondu la *politesse* avec la *déclaration* (56,3%). Le *maternel* (IDS) et la *déception* (DISA) sont bien reconnus et ne montrent pas de confusions particulières avec les autres affects sociaux.

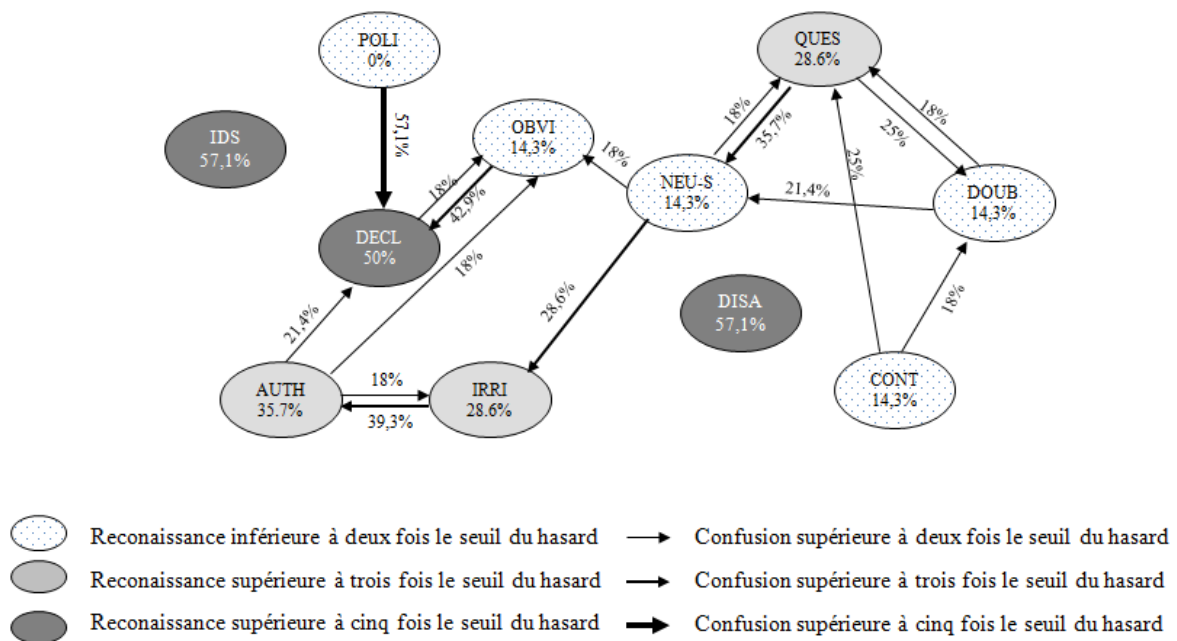


Figure 80 : Confusions des 11 affects sociaux chinois pour les 28 apprenants français de chinois mandarin dans la modalité « audio seul ». Les pourcentages d'identification sont indiqués dans les ovales et les taux de confusions sur les flèches. Seules les confusions qui affichent des scores supérieurs à deux fois le seuil du hasard sont rapportées.

En considérant toutes les confusions perçues entre affects sociaux, nous avons mené une analyse de classification hiérarchique afin d'identifier des classes perceptives de ces affects sociaux pour les apprenants français dans la condition « audio seul ».

Le dendrogramme de la figure 81 représente les résultats de cette analyse. Le graphique situé en haut à droite du dendrogramme illustre la perte d'inertie en fonction du nombre de groupes, à partir de laquelle on peut conclure qu'il est raisonnable de considérer cinq catégories d'affects sociaux.

Selon cette analyse, la première catégorie contient la *déclaration* (DECL), l'*évidence* (OBVI) et la *politesse* (POLI). L'*autorité* (AUTH) et l'*irritation* (IRRI), deux affects sociaux d'imposition sont regroupées dans la deuxième catégorie. La troisième catégorie contient le *mépris* (CONT), le *doute* (DOUB), l'exclamation de *surprise neutre* (NEU-S) et la *question* (QUES). La *déception* (DISA) et le *maternel* (IDS), affects sociaux bien reconnus par les sujets, forment des isolats puisqu'ils ne présentent pas de confusions particulières avec d'autres affects sociaux.

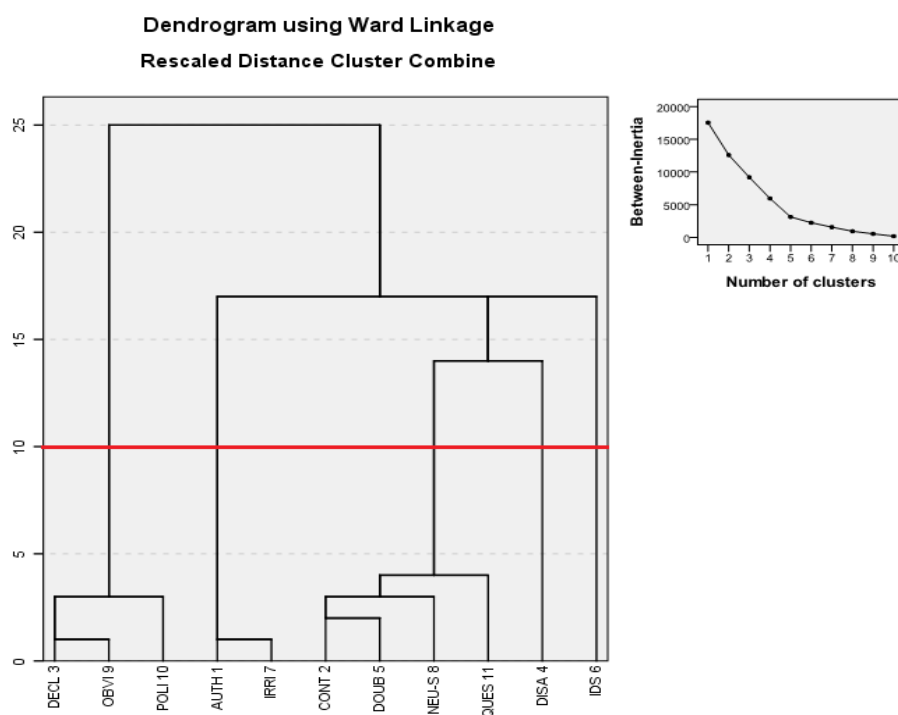


Figure 81 : Classification hiérarchique des 11 affects sociaux chinois pour les 28 apprenants français de chinois mandarin dans la modalité « audio seul ».

- Modalité « vidéo seul »

Les principales confusions entre les affects sociaux pour tous les sujets dans la modalité « vidéo seul » sont schématisées par la figure 82.

La *déclaration* (DECL) et la *déception* (DISA) sont bien perçues, alors que les indices visuels ne permettent pas une reconnaissance efficace, dans notre cas, de la *question* (QUES) et de l'exclamation de *surprise neutre* (NEU-S) (le taux d'identification est inférieur au hasard).

Des taux de confusion plus importants qu'en condition audio seul sont observés dans cette condition expérimentale. Le *maternel* (IDS), la *politesse* (POLI), l'*évidence* (OBVI) et la *question* (QUES) sont confondus avec la *déclaration* (DECL), elle-même n'étant confondue avec aucun autre affect social. Le *maternel* montre aussi des reports vers la *politesse*, qui est confondue avec l'*évidence*. L'*autorité* (AUTH) n'est confondue qu'avec l'*évidence*. La *déception* (DISA) attire des confusions provenant de l'*évidence*, de l'exclamation de *surprise neutre* (NEU-S), du *mépris* (CONT) et de l'*irritation* (IRRI), mais elle ne se reporte que sur le *doute* (DOUB). La *question* entretient d'importantes confusions avec la *déclaration* (46,4%) dans la condition visuelle seule. Le même constat

a été fait dans les résultats des sujets chinois pour les mêmes affects sociaux, ainsi que dans les résultats de la reconnaissance des affects sociaux multimodaux du français auprès de sujets français et d'apprenants chinois. Ce qui confirme l'hypothèse que la *déclaration* et la *question* constituent les deux modalités de base de l'énonciation, servant respectivement à donner ou à demander une information ; elles sont considérées comme affectivement neutres. Par conséquent, lorsque la locutrice exprime une déclaration et une question, elle ne manifeste pas de gestualité corporelle ni d'expression faciale spécifique, donc en l'absence d'information prosodique, les sujets ne parviennent pas à juger l'expression.

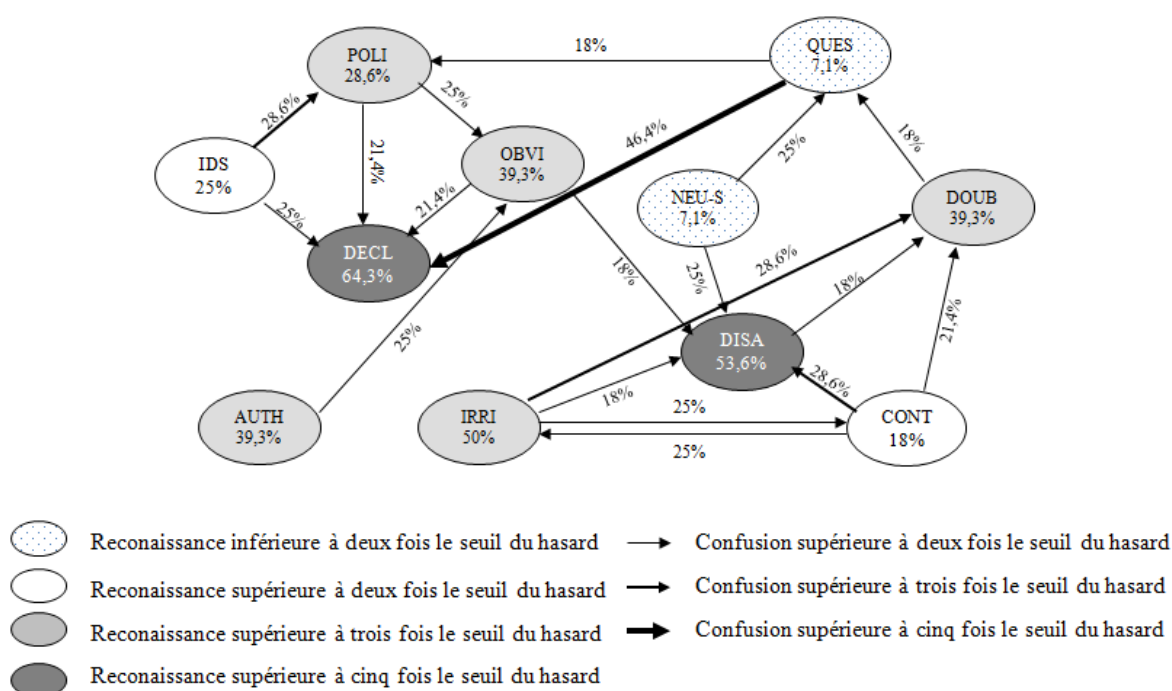


Figure 82 : Confusions des 11 affects sociaux chinois pour les 28 apprenants français de chinois mandarin dans la modalité « vidéo seul ». Les pourcentages d'identification sont indiqués dans les ovals et les taux de confusions sur les flèches. Seules les confusions qui affichent des scores supérieurs à deux fois le seuil du hasard sont rapportées.

En ce qui concerne le regroupement des affects sociaux sur la base de leur proximité dans la modalité visuelle seule, la figure 83 illustre les résultats de la classification hiérarchique. D'après la graphique de perte d'inertie en fonction du nombre de groupes, il est raisonnable de considérer trois catégories d'affects sociaux dans cette condition expérimentale.

Selon les résultats, la première catégorie regroupe la *déclaration* (DECL), la *politesse* (POLI), la *question* (QUES) et le *maternel* (IDS). La deuxième catégorie est composée de

l'exclamation de *surprise neutre* (NEU-S), de l'*évidence* (OBVI) et de l'*autorité* (AUTH). Enfin, la troisième catégorie contient le *mépris* (CONT), l'*irritation* (IRRI), le *doute* (DOUB) et la *déception* (DISA).

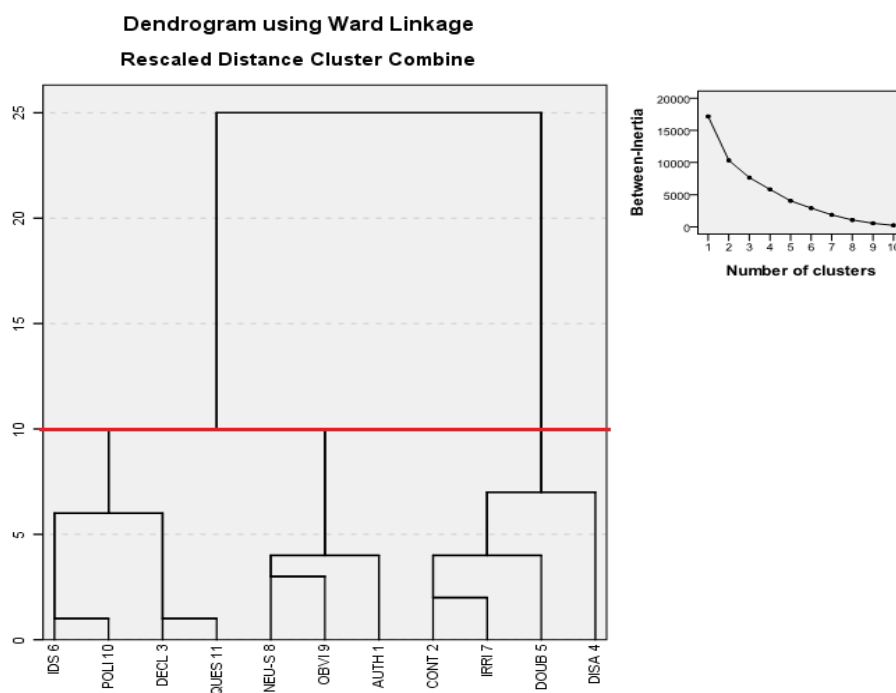


Figure 83 : Classification hiérarchique des 11 affects sociaux chinois pour les 28 apprenants français de chinois mandarin dans la modalité « vidéo seul ».

- Modalité « audio-vidéo »

La figure 84 représente les taux d'identification et les confusions entre affects sociaux dans la modalité audiovisuelle. Par rapport aux modalités « audio seul » et « vidéo seul », la modalité audiovisuelle véhicule plus d'informations expressives du fait de la synergie entre ces deux modalités. Par conséquent, les sujets reconnaissent mieux les affects sociaux dans cette condition expérimentale que dans les deux conditions séparées. On observe donc naturellement moins de confusions dans cette modalité. La plupart des affects sociaux sont bien reconnus dans cette modalité. La *déception* (DISA) reçoit le meilleur score de reconnaissance (85,7%).

La *question* (QUES) est confondue avec l'exclamation de *surprise neutre* (NEU-S) qui elle-même est confondue avec l'*irritation* (IRRI) ; l'*irritation* est confondue avec le *mépris* (CONT) qui se reporte principalement sur la *déception* (46,4%) ; la *politesse* (POLI) est

parfois confondue avec la *déclaration* (DECL). L'*autorité* (AUTH), le *doute* (DOUB), le *maternel* (IDS) et l'*évidence* (OBVI) sont bien identifiés sans confusions particulières.

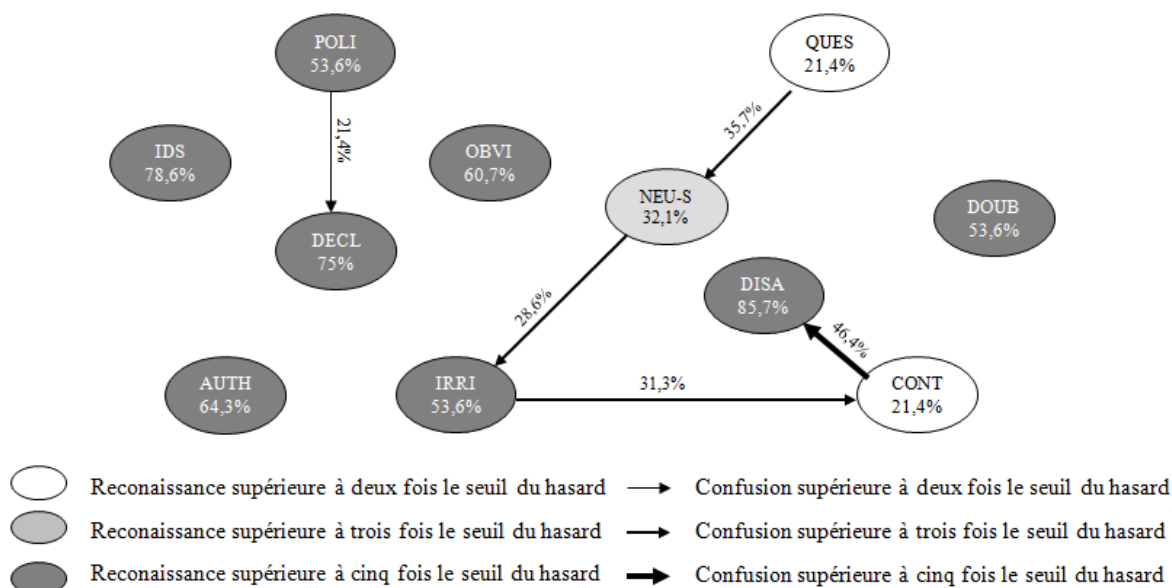


Figure 84 : Confusions des 11 affects sociaux chinois pour les 28 apprenants français de chinois mandarin dans la modalité « audio-vidéo ». Les pourcentages d'identification sont indiqués dans les ovales et les taux de confusions sur les flèches. Seules les confusions qui affichent des scores supérieurs à deux fois le seuil du hasard sont rapportées.

Les résultats de la catégorisation de ces affects sociaux dans la condition « audio-vidéo » sont illustrés sur la figure 85. D'après le graphique de perte d'inertie en fonction du nombre de groupes, les 11 affects sociaux se répartissent en cinq catégories pour cette condition expérimentale.

La première catégorie contient la *déclaration* (DECL) et la *politesse* (POLI), qui sont également confondues dans les deux modalités séparées ; la deuxième catégorie regroupe l'*autorité* (AUTH) et l'*évidence* (OBVI) ; le *mépris* (CONT) et la *déception* (DISA) sont classés dans la troisième catégorie ; la quatrième catégorie est composée de l'exclamation de *surprise neutre* (NEU-S), de la *question* (QUES), du *doute* (DOUB) et de l'*irritation* (IRRI). Ce regroupement de la *question*, du *doute* et de l'exclamation de *surprise neutre* est aussi observé dans la modalité « audio seul ». Le *maternel* (IDS), affect social bien reconnu, forme un isolat faute de confusions particulières.



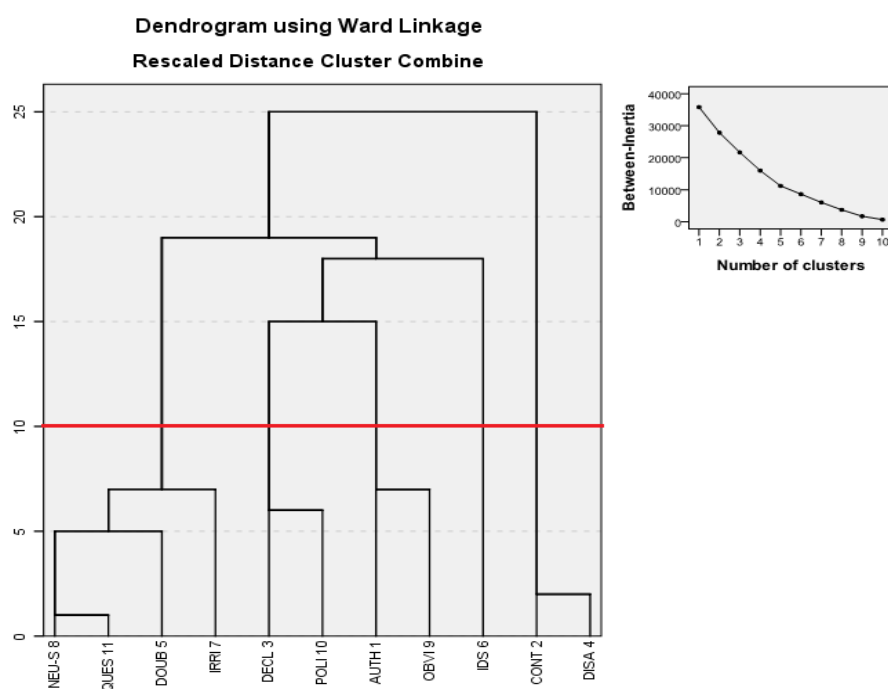
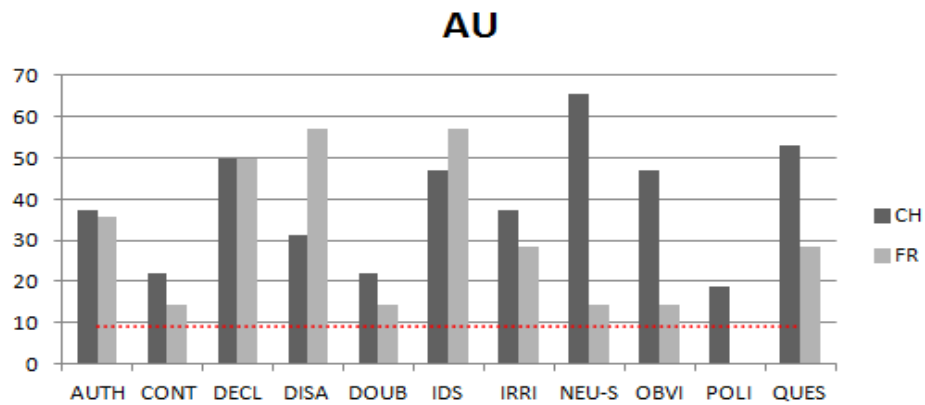


Figure 85 : Classification hiérarchique des 11 affects sociaux chinois pour les 28 apprenants français de chinois mandarin dans la modalité « audio-vidéo ».

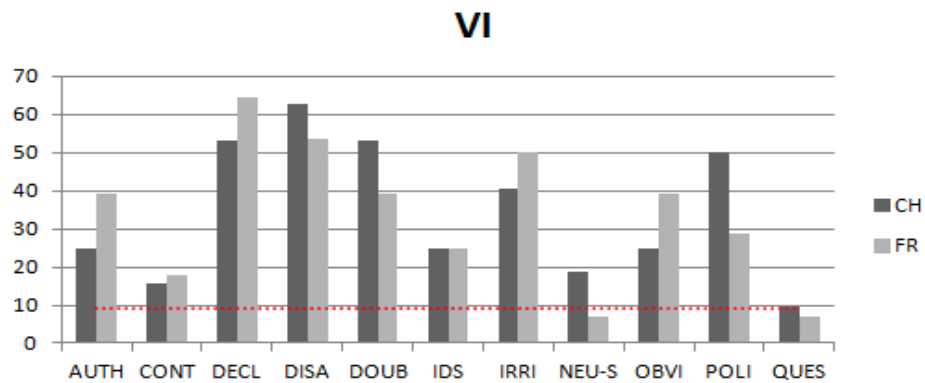
#### 4.6.5.4. Comparaison entre les apprenants français et les sujets natifs

Après avoir examiné la perception des affects sociaux audio-visuels du chinois par des apprenants français, nous tentons ici de comparer le comportement perceptif de ces apprenants avec celui des sujets chinois, que nous avons observé dans le chapitre 2 (voir partie 2.5).

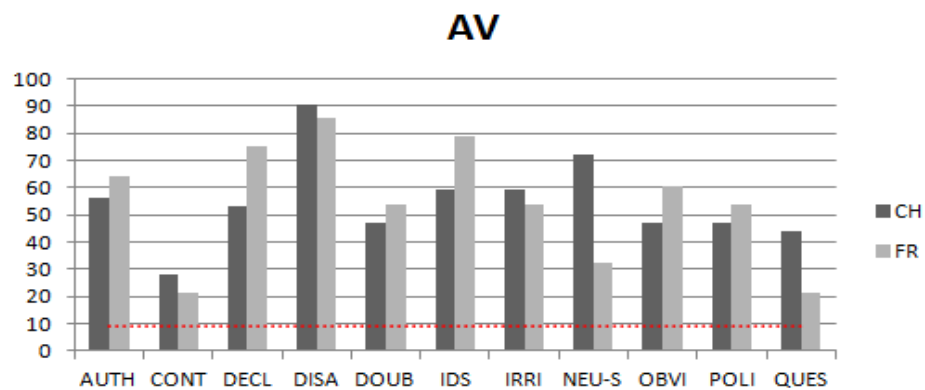
Au niveau du taux d'identification des affects sociaux dans les trois modalités, illustrés par la figure 86, nous pouvons constater que, dans la condition audio seule, les sujets natifs reconnaissent mieux l'exclamation de *surprise neutre* (NEU-S), l'*évidence* (OBVI), la *question* (QUES) et la *politesse* (POLI) que les apprenants français ; par contre, les apprenants français identifient mieux la *déception* que les sujets natifs (figure 86, a). Par ailleurs, la *politesse* n'est pas reconnue (0%) par les apprenants français dans cette modalité. Dans la condition visuelle seule (figure 86, b), les sujets natifs reconnaissent mieux la *politesse* ; les sujets français identifient mieux l'*évidence* et l'*autorité*, mais ils n'identifient pas l'exclamation de *surprise* et la *question*. Dans la condition audiovisuelle (figure 86, c), les sujets natifs perçoivent mieux l'exclamation de *surprise neutre* et la *question* ; et les apprenants français ont de meilleurs scores de reconnaissance pour la *déclaration* et le *maternel* (IDS).



(a) Modalité audio seule



(b) Modalité visuelle seule



(c) Modalité audio-visuelle

Figure 86 : Taux de reconnaissance des 11 affects sociaux chinois pour les sujets natifs et les apprenants français dans les trois modalités différentes : « audio seul » (haut) ; « vidéo seul » (milieu) ; « audio-vidéo » (bas).

En ce qui concerne l'effet de la modalité sur ces expressions socio-affectives pour les sujets natifs et les apprenants français de chinois, en général, ces expressions sont mieux reconnues par tous les sujets quand les informations auditive et visuelle sont mêlées. L'indice auditif et l'indice visuel ne jouent pas le même rôle pour les sujets natifs et non natifs lors de leur perception, sauf le *maternel*, qui est significativement mieux reconnu en « audio-vidéo » qu'en « vidéo seul » aussi bien par les sujets chinois que par les apprenants français ( $p < 0,05$ ). D'ailleurs, tous les deux groupes de sujets ont plus de facilité de reconnaître la *politesse* lorsque les indices visuels sont présents.

L'ordre de présentation des modalités a un effet significatif sur la reconnaissance des sujets natifs pour le *maternel* et l'exclamation de *surprise neutre* : ils sont significativement mieux reconnus dans l'ordre de « vidéo seul » - « audio seul » - « audio-vidéo ». Toutefois, ce facteur se révèle comme non significatif pour les apprenants français.

Enfin, en matière de confusions perceptives entre les affects sociaux, beaucoup de points communs sont remarqués entre les sujets français et chinois en fonction de la modalité. En « audio seul », l'*irritation* est confondue avec l'*autorité* ; la *politesse* est principalement mélangée avec la *déclaration* ; le *doute* et la *question* montrent des confusions réciproques, ainsi que la *déclaration* et l'*évidence* ; l'*autorité* se reporte parfois sur l'*évidence* et le mépris sur le *doute*. En « vidéo seul », l'*irritation* est confondue avec la *déception*, le *maternel* avec la *politesse*, le *mépris* avec la *déception*, ainsi que l'*autorité* avec l'*évidence*. La *question*, attitude très mal reconnue en vidéo seul aussi bien par les sujets chinois que par les apprenants français, est largement confondue avec la *déclaration* dans cette modalité. Le *doute* et l'exclamation de *surprise neutre* montrent des confusions avec la *question*. En « audio-vidéo », la *politesse* est confondue avec la *déclaration* par tous les sujets, ainsi que l'*irritation* avec le *mépris*.

#### 4.6.6. Conclusion

Dans cette étude de la perception multimodale des affects sociaux chinois, les mêmes stimuli utilisés lors du test perceptif pour des sujets chinois ont été retenus. Ces 33 stimuli ont été jugés par 28 apprenants français de chinois mandarin de niveau débutant (niveau A) et intermédiaire (niveau B).

Dans un premier temps, nous avons vérifié l'importance des facteurs sur la réponse des sujets grâce à une analyse de régression logistique mixte. Selon le résultat, les facteurs significatifs sont l'affect social et la modalité ; l'interaction entre l'affect social et la modalité se révèle également significative. L'ordre de présentation des modalités et le niveau de chinois des apprenants français ne présentent pas d'effets significatifs sur les réponses des sujets.

Ensuite, pour chaque affect social, des analyses de comparaison multiple ont été appliquées pour mettre en évidence les différences significatives de proportions de bonne réponse entre chacune des trois modalités. Selon les résultats, la *déception* est significativement mieux reconnue en « audio-vidéo » qu'en « vidéo seul » ( $p < 0,05$ ) ; pour le *doute*, la progression du taux de reconnaissance est très significative entre « audio seul » et « audio-vidéo » ( $p < 0,01$ ) ; l'*évidence* est significativement mieux identifiée en modalité audiovisuelle qu'en modalité audio seule ( $p < 0,01$ ) ; pour le *maternel*, les taux de reconnaissance en modalité audiovisuelle et audio seule sont significativement supérieurs à celui observé en modalité visuelle seule (en « audio-vidéo » :  $p < 0,01$  ; en « audio seul » :  $p < 0,05$ ), ce qui indique le rôle proéminent de l'information auditive dans l'expression.

Par la suite, une analyse de confusions perceptives entre les affects sociaux dans chaque condition expérimentale, basée sur la matrice de confusion, a été réalisée pour mettre en évidence le comportement perceptif des apprenants français dans chacune des modalités. Selon les résultats, la *politesse* est confondue avec la *déclaration* dans toutes les modalités ; la *question* est confondue avec l'exclamation de *surprise neutre* dans la condition audio seule et audiovisuelle, mais avec la *déclaration* et la *politesse* dans la condition visuelle seule ; le *mépris* montre des confusions vers la *déception* dans la condition visuelle seule et audiovisuelle, mais vers le *doute* et la *question* dans la condition audio seule. Signalons aussi que la *politesse* n'est pas reconnue par les apprenants français dans la modalité audio seule (0%), et est largement confondue avec la *déclaration* (57,1%). La *question* n'est pas reconnue dans la condition visuelle seule (7,1%), et est principalement confondue avec la *déclaration* (46,4%). Les mêmes observations ont été faites dans les réponses données par les sujets natifs (voir partie 2.5.6.3).

Au point de vue de la catégorisation de ces affects sociaux, le regroupement de la *déclaration* avec la *politesse* est observé dans les trois modalités ; le *doute*, la *question* et l'exclamation de *surprise neutre* sont réunis dans la même catégorie en « audio seul » et « audio-vidéo », mais ils sont répartis dans différents groupes en « vidéo seul » : la *question* est classée avec la *déclaration*, la *politesse* et le *maternel*, et le *doute* avec la *déception*, l'*irritation* et le *mépris* ; l'exclamation de *surprise neutre* est regroupé avec l'*autorité* et l'*irritation*. Le même constat a été remarqué dans l'étude de la perception multimodale des affects sociaux français auprès de sujets natifs et d'apprenants chinois (voir parties 4.4.4.3 et 4.5.5.3). Il est supposé que ces expressions interrogative (*question*), inattendue (exclamation de *surprise neutre*) et incertaine (*doute*) ont plus de similarités entre elles sur le plan prosodique que sur le plan posturo-mimo-gestuel.

Enfin, en comparant les comportements perceptifs des apprenants français avec les sujets natifs pour les mêmes stimuli, nous constatons que les sujets natifs et les apprenants français confondent certains affects sociaux de manière similaire dans chacune des modalités. D'un autre côté, la modalité ne joue pas le même rôle dans la reconnaissance des affects sociaux pour les deux groupes de sujets, par exemple les sujets chinois ont plus recours à l'information auditive pour identifier l'exclamation de *surprise neutre* ; les apprenants français reconnaissent mieux le *maternel* lorsque les indices auditifs sont présents. Tous les deux groupes de sujets reconnaissent la *politesse* plus à l'aide de l'information visuelle que de l'information auditive.

#### **4.7. Conclusion du chapitre**

Dans les chapitres précédents, nous avons examiné successivement la perception auditive et multimodale des expressions socio-affectives du chinois par des sujets natifs ; la perception interculturelle de la prosodie des affects sociaux chinois par des sujets français et vietnamiens ; et le traitement cognitif des affects sociaux chez des sujets chinois et français. Les résultats de ces études nous ont servi de référence pour cette étude de la perception multimodale des affects sociaux français par des apprenants chinois, et sur la perception multimodale des affects sociaux chinois par des apprenants français. Cette étude constitue une première étape des recherches spécifiques orientées vers la conception et l'application de méthodes d'enseignement de la prosodie socio-affective en français et en chinois langue étrangère ou seconde.

Après avoir conçu un corpus audiovisuel des affects sociaux français, en cohérence avec le corpus en chinois, nous avons réalisé, au sein de ce dernier chapitre de la thèse, une série d'expériences perceptives pour les expressions socio-affectives audiovisuelles en français et en chinois.

Dans le premier test perceptif, onze affects sociaux français véhiculés par un seul énoncé de quatre syllabes (« un jour deux nuits ») ont été jugés par 14 sujets français en multimodalité (audio seul, vidéo seul et audio-vidéo).

Le résultat d'une analyse de régression logistique mixte montre que les facteurs fixés, de même que leurs interactions d'ordre 2, sont significatifs à la reconnaissance de certains affects sociaux. Ces facteurs sont l'affect social, la modalité et l'ordre de présentation des modalités.

Une autre analyse de comparaisons multiples permet de mieux comprendre les effets de ces facteurs sur la reconnaissance des affects sociaux : l'information visuelle joue un rôle prépondérant dans la reconnaissance de l'expression de *mépris*, dont le taux d'identification en « audio seul » est significativement inférieur à ceux obtenus en « vidéo seul » et en « audio-vidéo » ( $p < 0,05$ ). Inversement, la prosodie contribue de façon majeure à la reconnaissance du *maternel*, affect social significativement mieux reconnu en audio seul et en audio-vidéo qu'en vidéo seul ( $p < 0,01$ ), et à la reconnaissance de la *question*, qui n'est pas reconnue par les sujets en vidéo seul, mais correctement reconnue en audio seul et en audio-vidéo. Le reste des affects sociaux sont reconnus plutôt grâce à la synergie entre les deux modalités.

En ce qui concerne l'effet de l'ordre de présentation des modalités, la majorité des affects sociaux ne montre pas de différences significatives entre les deux ordres de présentation des modalités, à l'exception du *mépris*, du *doute* et de l'*irritation* ( $p < 0,05$ ).

Par la suite, une analyse de confusions perçues entre les affects sociaux dans chaque condition expérimentale a été menée pour observer le regroupement des affects sociaux en fonction de modalité. En audio seul, la *déception*, le *maternel* et la *déclaration* sont bien reconnus et forment des isolats sans confusions particulières avec les autres affects sociaux, qui sont répartis en quatre catégories. En vidéo seul, le *mépris* est bien reconnu sans confusions particulières et les autres affects sociaux sont regroupés en quatre catégories différentes. En audio-vidéo, la *déclaration*, le *mépris*, la *déception* et l'*évidence*

sont bien identifiés par les sujets et les autres affects sociaux sont classés en trois catégories. Parallèlement, quelques observations intéressantes dans les trois modalités sont aussi à noter : l'*irritation* se rapproche de l'*autorité* dans les trois modalités ; la *question*, attitude non reconnue par les sujets natifs en vidéo seul, est largement confondue avec la *déclaration*; la *déception* est bien identifiée dans les trois modalités sans confusions particulières avec les autres affects sociaux ; la *déclaration* en « audio-vidéo » est reconnue par l'ensemble de sujets.

Dans le deuxième test perceptif, les mêmes stimuli utilisés dans le test précédent ont été retenus pour la perception de ces affects sociaux français par des apprenants chinois.

Selon le résultat d'une analyse de régression logistique mixte, sont significatifs les facteurs d'affect social, de modalité, d'ordre de présentation des modalités, de niveau de français des apprenants, ainsi que les interactions entre l'affect social et la modalité, et entre l'ordre de présentation des modalités et le niveau de français des apprenants.

Ensuite, le résultat de l'analyse de comparaisons multiples montre que l'information visuelle joue un rôle majeur pour la reconnaissance des expressions de *mépris* et d'*évidence*, qui sont significativement mieux reconnues en vidéo seul et en audio-vidéo qu'en audio seul ( $p < 0,01$ ) ; par contre, la prosodie contribue de façon majeure à la reconnaissance du *maternel*, affect social significativement mieux reconnu en audio seul et en audio-vidéo qu'en vidéo seul ( $p < 0,01$ ), et à celle de la *question*, attitude non reconnue en vidéo seul, mais est correctement reconnue en audio seul et en audio-vidéo. Le reste des affects sociaux sont reconnus plutôt grâce à la synergie entre les deux modalités, notamment le *doute*, qui est significativement mieux reconnu en audio-vidéo qu'en audio seul et vidéo seul ( $p < 0,05$ ).

Concernant l'effet de l'ordre de présentation des modalités, son interaction avec le niveau de français des apprenants chinois a un effet significatif sur la reconnaissance des affects sociaux. Cependant, il ne s'agit que de l'ordre de « vidéo seul » - « audio seul » - « audio-vidéo », dans lequel les apprenants de niveau intermédiaire identifient significativement mieux les expressions socio-affectives que les apprenants de niveau débutant.

Enfin, une analyse de confusions perçues entre les affects sociaux dans chaque condition expérimentale a été menée pour observer le regroupement des affects sociaux en

fonction de la modalité. En audio seul, la *déception* et le *maternel* sont bien reconnus et sans confusions particulières avec les autres affects sociaux qui sont répartis en trois catégories. En vidéo seul, les onze affects sociaux sont classés en quatre catégories. En audio-vidéo, la *déclaration* est bien identifiée par les sujets et les autres affects sociaux sont classés en quatre catégories. Parallèlement, l'*autorité* se reporte sur l'*irritation* dans les trois modalités, ainsi que la *politesse* sur le *maternel* et le *doute* sur la *question* ; la *question* et l'exclamation de *surprise neutre* ne sont pas reconnues par les apprenants en vidéo seul, et sont largement confondues avec la *déclaration*.

Dans le troisième test perceptif, les mêmes stimuli utilisés dans l'expérience multimodale de onze affects sociaux chinois perçus par des sujets natifs (voir partie 2.5) ont été retenus pour la perception de ces affects sociaux chinois par des apprenants français.

Tout d'abord, selon le résultat de l'analyse de régression logistique, les facteurs d'affect social et de modalité se révèlent comme significatifs, ainsi que leur interaction, par contre, l'ordre de présentation des modalités et le niveau de chinois des apprenants français ne le sont pas.

Le résultat de l'analyse de comparaisons multiples montre que la prosodie joue un rôle majeur dans la reconnaissance du *maternel*, qui est significativement mieux reconnu en audio seul ( $p < 0,05$ ) et en audio-vidéo ( $p < 0,01$ ) qu'en vidéo seul. Le *doute* et l'*évidence* montrent un taux d'identification significativement plus élevé en audio-visuel qu'en audio seul. Ce qui implique une contribution importante de la modalité visuelle à ces expressions. Le reste des affects sociaux sont reconnus plutôt grâce à la synergie entre les deux modalités.

Ensuite, à travers une analyse de confusions perçues entre affects sociaux dans chaque condition expérimentale, nous pouvons observer le regroupement des affects sociaux en fonction de la modalité. En audio seul, la *déception* et le *maternel* sont bien reconnus et sans confusions particulières. Les autres affects sociaux sont classés en trois catégories. En vidéo seul, la *déception* ne montre pas de confusions particulières et les autres affects sociaux sont répartis en trois catégories. En audio-vidéo, l'*irritation* et le *maternel* sont bien identifiés par les sujets et le reste des affects sociaux sont classés en quatre catégories.



Par ailleurs, quelques confusions importantes sont aussi observées: la *politesse* est confondue avec la *déclaration* dans toutes les modalités ; la *question* est confondue avec l'exclamation de *surprise neutre* dans la condition audio seule et audiovisuelle, mais avec la *déclaration* et la *politesse* en vidéo seul ; le *mépris* se reporte sur la *déception* en vidéo seul et en audio-vidéo, mais avec le *doute* et la *question* en audio seul. Il est aussi constaté que la *politesse* n'est pas reconnue par les apprenants français dans la modalité audio seule, et est largement confondue avec la *déclaration*. La *question* et l'exclamation de *surprise neutre* ne sont pas reconnues en vidéo seul, et la *question* est particulièrement confondue avec la *déclaration*.

Pour conclure, en confrontant les résultats de ces expériences multimodales des affects sociaux français et chinois par des sujets natifs et des apprenants étrangers, nous pouvons souligner quelques affects sociaux qui se montrent particuliers lors de la perception des sujets. Il s'agit de l'*évidence*, de la *politesse*, du *maternel*, de l'*autorité*, de la *question* et du *mépris*.

Pour l'*évidence* en audio seul, lorsqu'elle est exprimée en français, les sujets chinois la jugent proche de l'exclamation de *surprise neutre* et du *doute* ; lorsqu'elle est exprimée en chinois, ils la considèrent proche de l'*autorité*, ce qui est cohérent avec leur proximité conceptuelle chez les chinois. En écoutant attentivement le stimulus auditif de l'*évidence* en français, il apparaît que la locutrice utilise une intonation montante à la fin de l'énoncé, ce qui est contraire à l'expression en chinois qui se réalise par une intonation descendante. Cela pourrait être une explication possible à ce décalage perceptif de l'*évidence* chez les sujets chinois.

Pour l'expression prosodique de *politesse* chinoise, les apprenants français comme les sujets chinois la confondent largement avec la *déclaration* en l'absence d'information visuelle. Cependant, ce n'est pas le cas de la *politesse* en français. Une suggestion pourrait en être déduite : l'expression de *politesse* en chinois est moins marquée au niveau prosodique qu'au niveau visuel, et en tant qu'affect étroitement lié à la relation sociale interpersonnelle, son expression a plus recours aux indices posturo-mimo-gestuelles transmises dynamiquement dans l'interaction face-à-face.

En ce qui concerne le *maternel*, l'*autorité*, la *question* et le *mépris*, qu'ils soient exprimés en chinois ou en français, tous les sujets les perçoivent de manière similaire, quelle que soit leur langue maternelle.

Pour le *maternel* en vidéo seul, tous les sujets le considèrent proche de la *politesse* ; par contre en audio seul, ils l'identifient bien sans confusions particulières avec les autres affects sociaux. Ainsi, le *maternel* est plutôt un affect social qui s'identifie par ses caractéristiques acoustiques.

Pour l'*autorité*, en français comme en chinois, tous les sujets la perçoivent acoustiquement proche de l'*irritation*, par contre visuellement proche de l'*évidence*. Au niveau conceptuel, elle est jugée proche de l'*évidence* par tous les sujets quelle que soit leur langue maternelle. Ces constats indiquent qu'en prosodie, l'attitude autoritaire ou d'imposition ressemble à l'attitude irritée ou d'impolitesse dans une certaine mesure, mais à l'*évidence* quand les seuls indices visuels se présentent. Cela implique une similarité prosodique entre l'*autorité* et l'*irritation* à la fois en chinois et en français, et la perception visuelle de l'*autorité* est plus cohérente avec son traitement conceptuel.

La *question* est largement confondue avec la *déclaration* en vidéo seul aussi bien par les sujets chinois que par les sujets français. Cela pourrait être dû au fait que la *déclaration* et la *question* constituent les deux modalités de base de l'énonciation, cherchant respectivement à donner ou à demander une information ; elles sont considérées comme affectivement neutres. Par conséquent, lorsque la locutrice exprime une déclaration et une question, elle ne manifeste pas de gestualité corporelle ni d'expression faciale spécifique, donc en l'absence d'information prosodique, les sujets ne parviennent pas à juger l'expression.

Enfin, pour le *mépris*, quelques différences sont remarquées en fonction de la langue : en audio seul, lorsqu'il est exprimé en chinois, il est confondu avec le *doute* et la *question* par tous les sujets français et chinois ; lorsqu'il est exprimé en français, les apprenants chinois le jugent proche du *doute* et de la *question*, comme leur perception du *mépris* en chinois, alors que les sujets français natifs le considèrent proche de la *déclaration* et de la *politesse*.

## Conclusions et perspectives

Dans ces travaux de thèse, nous avons mené une série d'expériences perceptives sur les affects sociaux du chinois et du français dans le but d'atteindre quatre objectifs principaux : (1) observer la perception et les confusions de la prosodie des affects sociaux chinois aussi bien dans un cadre intra-culturel que dans un cadre interculturel ; (2) examiner le traitement cognitif des affects sociaux chez des sujets chinois et français ; (3) mettre en évidence les caractéristiques prosodiques saillantes des expressions socio-affectives en chinois ; (4) comparer la perception multimodale des affects sociaux du français et du chinois respectivement par des apprenants chinois et français avec la perception des natifs. Cette thèse constitue une première étape de la recherche sur l'enseignement/apprentissage de la prosodie socio-affective du français et du chinois pour les apprenants en langue étrangère (LE) ou en langue seconde (L2).

Dans cette dernière partie, nous résumerons les principaux résultats obtenus dans les études effectuées pour réaliser ces objectifs et nous proposerons également des pistes pour les recherches à venir.

### **Conclusions**

D'après les précédents travaux sur les affects sociaux présentés dans notre état de l'art, nous reprenons trois hypothèses. Selon la première, les affects sociaux émergent de manière volontaire et intentionnelle chez le locuteur pendant la communication, alors que les émotions sont plutôt déclenchées par un contrôle involontaire. Ensuite, les affects sociaux sont composés, en plus des attitudes, des affects qui véhiculent des valeurs sociales hiérarchiques, qui définissent les rôles sociétaux des locuteurs, et qui décrivent le contexte social général où se déroule l'interaction (Aubergé, 2002a). Puisque les affects sociaux sont encodés conventionnellement dans une culture et une langue, ils devront être appris par les apprenants de LE et L2, si les valeurs d'affects sociaux ne se partagent pas entre la langue cible et la langue source. A partir de ces principes, nous avons choisi pour notre recherche 19 affects sociaux communs des cultures chinoise et française. Les affects sociaux sont représentés par des étiquettes

vernaculaires et définis par des exemples de situation communicative où chaque affect social peut se produire.

Par la suite, nous avons construit un large corpus audio-visuel des affects sociaux en chinois mandarin, dans lequel 152 énoncés sélectionnés en fonction de la longueur des phrases (nombre de syllabes), du placement des tons et de la structure syntaxique ont été produits avec ces 19 affects sociaux par une enseignante de français, de langue maternelle chinoise et habituée à produire des énoncés réalistes pour un usage courant.

Ensuite, des études sur le terrain ont été menées selon ce cadre théorique et méthodologique élaboré puis développé dans les travaux d'Aubergé et al. (1997), Grépillat (1996), Morlec (1997) pour les attitudes françaises, de Diaféria (2002) pour l'anglais britannique, de Shochi (2008) pour le japonais et enfin de Mac (2012) pour le vietnamien.

En premier lieu, les analyses perceptives réalisées sur des auditeurs chinois ont validé globalement la qualité du corpus. Le taux d'identification de presque tous les affects sociaux était supérieur à celui du hasard, à l'exception de la *confiance*. La longueur des énoncés avait un effet significatif ( $p < 0,01$ ) sur la reconnaissance des affects sociaux : les énoncés monosyllabiques ont reçu les taux d'identification les plus bas et se différencient significativement des autres longueurs ( $p < 0,01$ ). Le genre avait aussi un effet significatif ( $p < 0,01$ ) mais de peu d'ampleur par rapport aux facteurs d'affect social et de longueur de l'énoncé. Sur la base des confusions perçues entre les affects sociaux, ils ont été classés en sept catégories, mais d'un point de vue plus général, les expressions d'exclamation de *surprise neutre, négative et positive*, de *doute* et de *question* ont été séparées de l'ensemble des autres affects sociaux. Cela montre la cohérence entre la perception et l'expression de ces affects sociaux, au travers desquelles la locutrice « interroge » son interlocuteur, notamment à l'aide d'une intonation montante en fin d'énoncé.

Ensuite, des analyses objectives des indices prosodiques pour les affects sociaux exprimés en chinois ont été réalisées en deux étapes. La première étape consistait à extraire les caractéristiques prosodiques saillantes de ces expressions socio-affectives à partir de dix indices prosodiques. La seconde étape ciblait les similarités (ou dissimilarités) des contours intonatifs de ces expressions en prenant en compte deux

indices que sont le RMS de F0 et la corrélation pondérée de la forme du contour. Ces analyses nous ont donné les informations suivantes :

Tout d'abord, les 19 affects sociaux pouvaient être réunis en cinq groupes selon les dix indices prosodiques examinés :

- *L'admiration*, le *maternel*, le *mépris* et le *doute* sont caractérisés par une durée moyenne plus longue de la première et de la dernière syllabe, et donc de l'ensemble de l'énoncé, de même que par des valeurs élevées de tous les indices de F0 ;
- *L'autorité*, la *déclaration*, la *résignation* et la *déception* sont au contraire caractérisées par une durée moyenne courte notamment pour la dernière syllabe, et également par un registre bas, une plage étroite de F0 et moins de variabilité de F0 ;
- *L'évidence*, la *séduction*, la *politesse* et la *confiance* sont réalisées avec des caractéristiques de F0 proches de celles du groupe précédent, mais elles présentent une durée plus importante.
- *L'ironie* et *l'intimité* se caractérisent par une durée moyenne très courte et donc une énonciation rapide, ainsi qu'une importante pente descendante de F0 sur l'ensemble de l'énoncé;
- La *surprise positive*, la *surprise négative*, la *surprise neutre*, la *question* et *l'irritation* sont réalisées avec une large plage de F0, un niveau élevé de F0 moyenne pour la première et la dernière syllabe ainsi que pour l'ensemble de l'énoncé, et également par une durée moyenne élevée de la dernière syllabe.

Par la suite, ces affects sociaux ont été répartis en différents groupes, selon la longueur des énoncés, et sur la base des mesures de RMS de F0 et de la corrélation de la forme du contour de F0. Cependant, au plus haut niveau des dendrogrammes, la majorité des affects ont été classés à l'identique en deux groupes principaux que ce soit en fonction du RMS ou de la corrélation de forme de contour, en dépit des variations de longueur d'énoncé. Ce regroupement correspondait à celui observé lors du test perceptif chez les sujets natifs.

D'autre part, les résultats de l'analyse acoustique nous ont permis de comprendre certaines des confusions d'affects sociaux perçues par les sujets natifs, comme la

confusion de la *déclaration* avec l'*évidence* et la *confiance*, celle entre l'*irritation* et l'*autorité*, et celle entre la *déception* et la *résignation*. Cependant, certaines confusions observées peuvent avoir des causes plus complexes et ne peuvent pas s'expliquer uniquement par des similarités en prosodie : c'est le cas de la confusion entre le *maternel* et la *séduction*.

Parallèlement, la perception multimodale des affects sociaux chinois a été examinée. Onze affects sociaux ont été choisis afin de maximiser la prototypicité, et à des fins d'usage didactique de la prosodie socio-affective du chinois et du français LE ou L2. Les résultats de l'analyse ont montré que la synergie des modalités auditive et visuelle permet aux sujets de mieux reconnaître les affects sociaux qu'une seule modalité dans la plupart des cas. La modalité des expressions socio-affectives et l'ordre de présentation des modalités avaient des effets significatifs sur la reconnaissance de certains affects sociaux. Précisément, les expressions d'exclamation de *surprise neutre* et de *question* sont plus dépendantes de l'information auditive alors que celle de *politesse* a été mieux reconnue lorsque les informations visuelles sont présentes ; les différentes modalités n'ont pas exercé d'influence particulière sur la reconnaissance de la *déclaration*, qui a reçu de bons scores de reconnaissance dans les trois conditions expérimentales ; enfin, les autres expressions socio-affectives ont eu plutôt recours aux informations auditives et visuelles combinées. Au niveau des confusions perceptives, les onze affects sociaux se sont regroupés dans des catégories différentes en fonction des modalités considérées. Il est à noter que la *question* a été classée avec la *déclaration* en modalité visuelle seule, mais avec le *doute* dans les modalités audio seule et audiovisuelle.

En deuxième lieu, pour pouvoir comparer les comportements perceptifs des sujets natifs avec celui des sujets français (puisque notre recherche se base sur une étude croisée du chinois et du français), et examiner l'effet potentiel du ton sur la reconnaissance des affects sociaux, nous avons mené une étude interculturelle de la perception de ces 19 affects sociaux chinois par des sujets français et vietnamiens. Attendu que leur proximité culturelle et linguistique avec les chinois est considérée plus forte que celle entre les français et les chinois, les sujets vietnamiens constituent un groupe de référence nécessaire pour analyser les résultats obtenus chez les sujets français.

Dans un premier temps, les résultats de l'analyse ont montré que l'avantage culturel ou linguistique des sujets vietnamiens (d'une langue tonale) par rapport aux sujets français (d'une langue non tonale) n'a pas facilité leur reconnaissance des affects sociaux chinois, et que les tons chinois avaient un effet significatif sur l'identification de certains affects sociaux pour tous les sujets. De plus, la variation du ton final de l'énoncé a plus influencé la reconnaissance des affects sociaux que celle du ton initial, lorsque les autres syllabes étaient fixées au ton plat (ton 1).

Quant aux confusions perçues entre les affects sociaux, la *déclaration* était l'attitude la plus attractive, - la plus choisie comme réponse dans l'expérience perceptive, chez les sujets français et vietnamiens comme chez les sujets natifs. La *question* ne s'est pas reportée particulièrement sur les autres affects sociaux, elle était cependant la cible attractive des transferts des trois expressions de *surprise* et du *doute*. L'*irritation* et l'*autorité* ont été essentiellement confondues l'une avec l'autre, ainsi que la *déception* et la *résignation*. L'*admiration* n'a pas montré de confusions particulières avec les autres affects sociaux.

Enfin, en matière de classement des affects sociaux, au plus haut niveau de différenciation, les affects sociaux ont été classés en deux catégories identiques pour les sujets français, vietnamiens et chinois. Dans le cadre d'une description plus fine, les sujets français et vietnamiens ont classé les affects sociaux en huit catégories, alors que les sujets chinois ne les ont classés qu'en sept catégories. La différence majeure de classification des affects sociaux entre les sujets natifs et non natifs repose essentiellement sur l'*admiration*, le *mépris*, l'*ironie* et l'*intimité*.

Les similarités mais aussi les décalages perceptifs des mêmes affects sociaux chinois entre les sujets chinois et les sujets français nécessitent de mettre au clair leur source potentielle. Ainsi, nous avons eu recours à une étude « conceptuelle » et psycho-acoustique reposant sur le postulat selon lequel les concepts des affects sociaux varieraient éventuellement d'une culture à l'autre et que ces variations seraient à l'origine des décalages perceptifs des affects sociaux entre deux cultures différentes. Pour vérifier cette hypothèse, nous avons effectué, d'une part, un test conceptuel pour les 19 affects sociaux chez des sujets chinois et français afin de mesurer leurs distances conceptuelles et de les comparer entre les deux cultures. D'autre part, nous avons réalisé

un test psycho-acoustique pour mesurer les distances acoustiques entre les affects sociaux et les comparer avec leurs distances conceptuelles au sein de la culture chinoise.

Le résultat de l'analyse de correspondance a indiqué que les variations de dispersion des affects sociaux étaient plus liées aux différences entre les affects sociaux eux-mêmes qu'à la langue/culture des sujets. Le genre des sujets n'avait pas un effet important sur la distribution des affects sociaux. Selon l'analyse de classification hiérarchique, la plupart des affects sociaux ont été classés de manière similaire par les sujets chinois et français à partir de leurs proximités, à l'exception de la *question*, qui a été jugée proche des affects sociaux positifs par les sujets chinois, mais proche des affects sociaux négatifs par les sujets français. Cela implique que le concept de la majorité des affects sociaux existe de manière similaire dans les cultures chinoise et française en dépit de leur distance culturelle.

D'un point de vue plus général, les concepts des affects sociaux ont été classés par tous les sujets, quelle que soit leur langue/culture, en quatre groupes en fonction de leur valence et de leur arousal :

- Les affects sociaux de valence positive et avec un fort arousal : l'exclamation de *surprise positive*, l'*admiration*, le *maternel*, l'*intimité*, la *politesse* et la *séduction* ;
- Les affects sociaux de valence positive et avec un faible arousal : l'*évidence*, la *confiance*, la *déclaration*, sauf l'*autorité* qui montre une valence positive et un arousal fort ;
- Les affects sociaux de valence négative et avec un fort arousal : l'exclamation de *surprise négative*, la *déception*, l'*ironie*, le *mépris* et l'*irritation* ;
- Les affects sociaux de valence négative et avec un faible arousal : la *question*, le *doute* et l'exclamation de *surprise neutre*.

En confrontant les distances conceptuelles avec les distances psycho-acoustiques des expressions socio-affectives chinoises, nous avons constaté que les affects sociaux ont été regroupés en deux groupes identiques : l'un contient les attitudes d'assertion et de soumission, ainsi que les affects révélant la relation interpersonnelle et le contexte interactif (*déclaration*, *confiance*, *évidence*, *admiration*, *résignation*, *déception*, *politesse*, *autorité*, *séduction*, *intimité* et *maternel*) ; l'autre est composé des attitudes interrogatives, inattendues et incertaines, ainsi que des attitudes de refus (*question*,



*surprise négative, surprise positive, surprise neutre, doute, mépris, ironie et irritation*). Signalons que la plupart des affects sociaux dans le second groupe ont montré une forte corrélation entre leur concept et leur prosodie. Inversement, la plupart des affects sociaux du premier groupe ont montré un certain décalage et ont été majoritairement corrélés avec un autre affect social à cause de similarités prosodiques. Il s'agit de la *déclaration*, de l'*autorité*, de la *politesse*, de l'*évidence*, de la *confiance*, de la *déception*, de la *résignation* et de la *séduction*.

Par ailleurs, dans les deux tests, la plupart des connaissances sur les affects sociaux ont été globalement partagées par tous les sujets, quels que soient leur contexte culturel et linguistique, leur genre ou la manière de présenter les affects sociaux (concepts ou entrées acoustiques).

En se reposant sur les résultats des tests conceptuel et acoustique, nous pouvons en conclure que les sujets chinois et les sujets français traitent les concepts d'affects sociaux d'une manière similaire ; les décalages perceptifs des affects sociaux chinois entre les sujets chinois et français proviennent dans une large mesure de leur façon de percevoir la prosodie socio-affective chinoise. Ainsi, par rapport au concept, la prosodie jouerait un rôle majeur dans la perception des expressions socio-affectives.

Enfin, dans le dernier chapitre notre intérêt a porté sur une étude de la perception multimodale des affects sociaux français par des apprenants chinois, et des affects sociaux chinois par des apprenants français. Cette étude s'est donnée comme objectif d'examiner la contribution des informations auditives et visuelles pour la perception des affects sociaux lors d'une interaction face-à-face et dynamique en langue étrangère, ainsi que d'évaluer l'effet de l'apprentissage de la langue cible sur la perception des affects sociaux des apprenants LE ou L2. A long terme, cette étude constitue une première étape à des recherches spécifiques orientées vers la conception et l'application de méthodes d'enseignement de la prosodie socio-affective en français et en chinois langue étrangère ou seconde.

Au préalable, nous avons effectué une série d'expériences perceptives pour une sélection de onze expressions socio-affectives audio-visuelles en français et en chinois,

perçues respectivement par des apprenants chinois et français de niveau débutant (niveau A) et intermédiaire (niveau B).

Premièrement, pour les apprenants chinois de français, la modalité des expressions socio-affectives (audio seul, vidéo seul ou audio-vidéo) a eu un effet significatif sur la reconnaissance de certains affects sociaux français. Précisément, l'information visuelle a joué un rôle majeur dans la reconnaissance des expressions de *mépris* et d'*évidence*, significativement mieux reconnues en vidéo seul et en audio-vidéo qu'en audio seul ( $p < 0,01$ ) ; au contraire, la prosodie a contribué de façon majeure à la reconnaissance du *maternel*, affect social significativement mieux reconnu en audio seul et en audio-vidéo qu'en vidéo seul ( $p < 0,01$ ), et à celle de *question*, qui n'a pas été identifiée par les sujets en vidéo seul, mais a été correctement reconnue en audio seul et en audio-vidéo. Le reste des affects sociaux ont été mieux reconnus grâce à la synergie des deux modalités, notamment le *doute*, qui a été significativement mieux reconnu en audio-vidéo qu'en audio ou vidéo seuls ( $p < 0,05$ ). Au niveau des confusions perceptives, l'*autorité* s'est reportée sur l'*irritation* dans les trois modalités, ainsi que la *politesse* sur le *maternel* et le *doute* sur la *question* ; la *question* et l'exclamation de *surprise neutre* n'ont pas été reconnues par les apprenants en vidéo seul, et ont été largement confondues avec la *déclaration*.

L'interaction entre l'ordre de présentation des modalités et le niveau des apprenants a montré un effet significatif sur la reconnaissance des affects sociaux. Lorsque les stimuli étaient présentés dans l'ordre de « vidéo seul » - « audio seul » - « audio-vidéo », les apprenants de niveau intermédiaire ont significativement mieux identifié les expressions socio-affectives que les apprenants de niveau débutant.

En comparant la perception des mêmes affects sociaux exprimés en français et en chinois par les sujets chinois (sujets chinois non exposés à la langue/culture française et apprenants chinois du français), l'attitude d'*évidence* en audio seul a attiré notre attention : lorsqu'elle est exprimée en français, les sujets chinois l'ont jugée proche de l'exclamation de *surprise neutre* et du *doute* ; par contre, lorsqu'elle est exprimée en chinois, ceux-ci l'ont considérée proche de l'*autorité*. Ces observations sont cohérentes avec leur proximité conceptuelle en chinois. Ce décalage perceptif de l'*évidence* chez les sujets chinois pourrait provenir de ses réalisations prosodiques différentes en français

et en chinois : une intonation plutôt montante à la fin de l'énoncé a été produite en français alors que l'intonation est plutôt descendante en chinois.

Pour les apprenants français du chinois, seul le facteur de modalité a eu un impact significatif sur la reconnaissance des affects sociaux chinois. Le niveau de chinois du sujet et l'ordre de présentation des modalités ont été testés non significatifs. Les résultats de l'analyse de comparaisons multiples ont montré que la prosodie a joué un rôle majeur dans la reconnaissance du *maternel*, significativement mieux reconnu en audio seul ( $p < 0,05$ ) et en audio-vidéo ( $p < 0,01$ ) qu'en vidéo seul. Le reste des affects sociaux a été mieux reconnu grâce à la synergie des deux modalités. Quant aux confusions perceptives, quelques confusions importantes ont été observées: la *politesse* a été confondue avec la *déclaration* dans toutes les modalités ; la *question* a été confondue avec l'exclamation de *surprise neutre* dans la condition audio seule et audiovisuelle, mais avec la *déclaration* et la *politesse* en vidéo seul ; le *mépris* s'est reporté sur la *déception* en vidéo seul et en audio-vidéo, mais sur le *doute* et la *question* en audio seul. La *politesse* n'a pas été reconnue par les apprenants français dans la modalité audio seule, mais a été largement confondue avec la *déclaration*. La *question* et l'exclamation de *surprise neutre* n'ont pas été reconnues en vidéo seul, et la *question* a été particulièrement confondue avec la *déclaration*.

En confrontant les résultats de reconnaissance des mêmes affects sociaux exprimés en chinois et en français pour les sujets français (sujets français non exposés à la langue/culture chinoise et apprenants français du chinois), nous avons constaté que les sujets français ont très largement confondue la prosodie de la *politesse* chinoise avec celle de la *déclaration* lorsque l'information visuelle est absente (comme les sujets chinois) ; cependant, ce n'est pas le cas de la *politesse* française. Il est supposé que l'expression de *politesse* en chinois est moins marquée au niveau prosodique qu'au niveau visuel ; l'expression de *politesse* est plus dépendante aux indices posturo-mimo-gestuelles transmis dynamiquement dans l'interaction face-à-face.

Mis à part les expressions d'*évidence* et de *politesse*, le *maternel*, l'*autorité*, la *question* et le *mépris*, méritent également notre attention, étant donné que tous les sujets, quelle que soit leur langue maternelle, les ont perçus de manière similaire, qu'ils soient exprimés en chinois ou en français.

- Le *maternel* : en vidéo seul, tous les sujets l'ont considéré proche de la *politesse* ; par contre en audio seul, ils l'ont bien identifié sans confusions particulières avec les autres affects sociaux. De ce fait, le *maternel* est plutôt un affect social qui s'identifie par ses caractéristiques acoustiques.
- L'*autorité* : en audio seul, tous les sujets l'ont perçue proche de l'*irritation* ; mais proche de l'*évidence* en vidéo seul. Au niveau conceptuel, elle a été considérée proche de l'*évidence* par tous les sujets chinois et français. Ces constats indiquent qu'en prosodie, l'attitude autoritaire ou d'imposition ressemble à l'attitude irritante ou d'impolitesse, mais elle est plus proche de l'*évidence* lorsque les seuls indices visuels se présentent. Par ailleurs, cette proximité prosodique de ces deux affects sociaux a déjà été observée dans l'analyse acoustique des affects sociaux chinois que nous avons présentée dans la partie 2.4. Ainsi, la similarité entre l'*autorité* et l'*irritation* en prosodie existe à la fois en chinois et en français, et la perception visuelle de l'*autorité* est plus cohérente avec son traitement conceptuel.
- La *question* a été largement confondue avec la *déclaration* en vidéo seul par tous les sujets chinois et français. Cette confusion pourrait être due au fait que la *déclaration* et la *question* constituent deux modalités de base de l'énonciation, servant respectivement à donner ou à demander une information ; elles sont considérées comme affectivement neutres. Par conséquent, lorsque la locutrice exprime une déclaration et une question, elle ne manifeste pas de gestualité corporelle ni d'expression faciale spécifique, et donc, en l'absence d'information prosodique, les sujets ne parviennent pas à juger l'expression.
- Le *mépris* : en audio seul, lorsqu'il a été exprimé en chinois, il a été confondu avec le *doute* et la *question* par tous les sujets français et chinois ; lorsqu'il a été exprimé en français, les apprenants chinois l'ont jugé proche du *doute* et de la *question*, comme leur perception auditive du *mépris* en chinois, alors que les sujets français natifs l'ont considéré proche de la *déclaration* et de la *politesse*.

## Perspectives

Dans le cadre des travaux de thèse, nous avons mené des études sur la perception de la prosodie des affects sociaux en chinois et en français vers une application pour l'apprentissage de ces deux langues chez des apprenants LE et L2. Ces premières études ouvrent également de nouvelles pistes de réflexions pour nos recherches futures.

Premièrement, étendre le corpus audio-visuel des affects sociaux chinois avec des énoncés de diverses longueurs et de structures syntaxiques plus variées sera utile afin d'étendre notre modèle prosodique selon la fonction syntaxique de la prosodie. De plus, il est aussi nécessaire de continuer à augmenter la couverture sociale et interactionnelle des affects sociaux.

D'autre part, au tout début de notre étude, nous avons enregistré, dans les mêmes conditions, les énoncés de notre corpus chinois avec un locuteur masculin de même origine régionale que notre locutrice. L'enregistrement de ce locuteur n'a pas été utilisé à cause de la complexité déjà importante de notre étude avec un seul locuteur. Cependant, valider ce deuxième locuteur ainsi qu'enregistrer plusieurs autres locuteurs est important pour examiner les effets du locuteur sur la production/perception des affects sociaux : par exemple, quel serait l'effet du genre, de l'âge, du niveau socioculturel et/ou régional ?

Nous avons également essayé dans notre thèse d'extraire quelques caractéristiques acoustiques globales des affects sociaux chinois sur la base des trois paramètres prosodiques que sont la fréquence fondamentale, l'intensité et la durée. Néanmoins, puisque la qualité vocale constitue un aspect important de l'expression affective (Campbell & Mokhtari, 2003), il sera pertinent d'examiner les caractéristiques physiques de la voix lors de la production de l'expression socio-affective. D'ailleurs, ces caractéristiques vocales nous aideront éventuellement à mieux comprendre certaines confusions perceptives comme celle entre le *maternel* et la *séduction*.

De plus, dans notre étude interculturelle de la perception de la prosodie des affects sociaux chinois, nous avons pu seulement comparer le comportement perceptif des sujets chinois avec celui des sujets français et vietnamiens. Pour chercher des affects sociaux universels, nous devons enrichir en langues ce paradigme d'étude interculturelle, telles que l'anglais, l'allemand, l'italien, le japonais ou le coréen.

Enfin, en vue de notre objectif final de la thèse, nous avons réalisé une série d'expériences perceptives sur les affects sociaux français et chinois chez des apprenants chinois et français en LE ou L2. Nous avons pris en compte l'effet potentiel de la modalité, de l'ordre de présentation des modalités et du degré d'acquisition de la langue cible sur la reconnaissance des affects sociaux. L'un des défauts de cette étude repose sur le choix du niveau de langue cible de l'apprenant, car nous n'avons pu recruter que des apprenants de niveau débutant et intermédiaire pour les expériences. En parallèle, le nombre des apprenants n'est sans doute pas suffisant pour conclure sur les tendances observées. Ainsi, ce travail constitue davantage une étude pilote, la poursuivre et recruter davantage d'apprenants à des niveaux variables (débutant, intermédiaire et avancé) permettrait de préciser nos premiers résultats.

D'autre part, notre travail s'est limité à la production d'un corpus de référence et à la perception d'affects sociaux « actés » : ce sont des matériaux qui devraient permettre une approche didactique de l'enseignement des prototypes prosodiques. Ainsi, nous tenterons de proposer des méthodes pour faciliter l'enseignement/apprentissage de la prosodie socio-affective en français et/ou en chinois. En s'inspirant des méthodes de correction phonétique en langue étrangère et seconde, nous envisagerons une méthode adaptée pour la « correction prosodique » des affects sociaux. Cet enseignement pourrait se faire soit à l'aide de descriptions de la production et de dessins de contextualisation socio-relationnelle des affects, ainsi que de jeux de rôle, ou soit à l'aide de logiciels spécifiques qui pourraient se baser sur les corpus que nous avons enregistrés.

Par la suite, la collection d'énoncés prosodiques spontanés des affects sociaux devra aussi être envisagée. En reproduisant le protocole de recueil de corpus spontané proposé par Antunes et al (2014), nous avons collecté, pour 12 locuteurs chinois d'origine régionale différente et âgés de 25 à 35 ans, des énoncés spontanés de confiance et de doute en chinois mandarin, grâce à un magicien d'Oz simulant un entretien d'embauche (en caméra cachée sans que les sujets ne connaissent le but de l'expérience). Les contenus linguistiques des énoncés sont comparables entre les locuteurs, et ces énoncés ont été également produits en condition actée par les mêmes sujets afin de les comparer aux énoncés spontanés. Nous envisageons des analyses approfondies sur ces données pour mettre en lumière des informations utiles à la compréhension des affects sociaux dans nos projets de recherches à venir.

# Publications

Lu, Y., Aubergé, V., & Rilliard, A. (2012). *Do you hear my attitude? Prosodic perception of social affects in Mandarin*. In proceedings of Speech Prosody 2012, 685-688, Shanghai, Chine.

Lu, Y., Aubergé, V., & Rilliard, A. (2012). *Tonal Influences on the prosodic Cross-linguistic Perception of Mandarin Social Affects by French and Vietnamese listeners*. In proceedings of the Third International Symposium of Tonal Aspect of Language, Nanjing, Chine.

Lu, Y., Aubergé, V., Rilliard, A., & Gu, W. (2012). *Prosodic Cross-linguistic Perception of Social Affects in Mandarin Chinese by natives, French and Vietnamese listeners*. In proceedings of GSCP 2012, 141-145, Belo Horizonte, Brésil.

Lu, Y., Aubergé, V., & Rilliard, A. (2012). *Entends-tu mes attitudes? Perception de la prosodie des affects sociaux en chinois Mandarin*. In proceedings of JEP 2012, 25-32, Grenoble, France.

Lu, Y., Aubergé, V., & Rilliard, A. (2013). *Cognitive distance of attitudes in Chinese*. In proceedings of WASSS 2013, Grenoble, France.

Lu, Y., Aubergé, V., Rilliard, A., & Gu, W. (2013). Perceptual study for Mandarin attitudinal speech. *Journal of School of Chinese Language and Culture*, 3, 169-174, Nanjing Normal University. (en chinois)

Lu, Y., Aubergé, V., Audibert, N., & Rilliard, A. (2014). *Audiovisual Perception of Expressions of Mandarin Chinese Social Affects by French L2 Learners*. In proceedings of Speech Prosody 2014, 169-173, Dublin, Ireland.

Lu, Y., Aubergé, V., & Rilliard, A. (2014). *Prosodic Profiles of Social Affects in Mandarin Chinese*. In proceedings of Speech Prosody 2014, 125-129, Dublin, Ireland.

# Bibliographie

- Abelin, Å., & Allwood, J. (2000). *Cross linguistic interpretation of emotional prosody*. In proceedings of the ISCA Tutorial and Research Workshop (ITRW) on Speech and Emotion, Newcastle, Northern Ireland.
- Abercrombie, D. (1967). *Elements of general phonetics*. Edinburgh: Edinburgh University Press.
- Alku, P., Bäckström, T., & Vilkman, E. (2002). Normalized amplitude quotient for parametrization of the glottal flow. *Journal of the Acoustical Society of America*, 112(2), 701-710.
- Allport, G. W. (1935). Attitudes. In C. Murchison (Ed.), *Handbook of social psychology* (pp. 798-844). Worcester, Mass: Clark University Press.
- Antunes, L., Aubergé, V., & Sasa, Y. (2014). *Certainty and uncertainty in Brazilian Portuguese: methodology of spontaneous corpus collection and data analysis*. In proceedings of the Speech Prosody, Dublin, Irlande.
- Arndt, H., & Janney, R. W. (1991). Verbal, prosodic, and kinesic emotive contrasts in speech. *Journal of pragmatics*, 15(6), 521-549.
- Arnold, M. B. (1960). *Emotion and Personality*: New York: Columbia University Press.
- Aubergé, V. (1991). *La synthèse de la parole : "des règles aux lexiques"*. (PhD Thesis), Grenoble, Université Pierre Mendès-France.
- Aubergé, V. (1992). Developing a structured lexicon for synthesis of prosody. In G. Bailly & C. Benoît (Eds.), *Talking Machines: Theories, Models and Designs* (pp. 307-321).
- Aubergé, V. (2002a). *A Gestalt Morphology of Prosody Directed by Functions*. In proceedings of the Speech Prosody, Aix-en-Provence.
- Aubergé, V. (2002b). Prosodie et émotion. *Actes des deuxièmes assises nationales du GdR I*, 3, 263-273.
- Aubergé, V. (2003). Expressions, attitudes et expressivité: une architecture cognitive distribuée pour les voies parlées des émotions. *Interfaces Prosodiques*.
- Aubergé, V. (2012). *Attitude vs. emotion: a question of voluntary vs. involuntary control*. In proceedings of the International Conference on Speech and Copora (GSCP2012), Belo Horizonte, Brazil.
- Aubergé, V., Grepillat, T., & Rilliard, A. (1997). *Can we perceive attitudes before the end of sentences? The gating paradigm for prosodic contours*. In proceedings of the Fifth European Conference on Speech Communication and Technology.
- Aubergé, V., & Rilliard, A. (2006). *More than pointing with the prosodic focus : the valence-intensity-domain (VID) model*. In proceedings of the 3rd International Conference on Speech Prosody, Dresden, Allemagne



- Audibert, N., Aubergé, V., & Rilliard, A. (2008). Emotions actées vs. spontanées : variabilité des compétences perceptives. In proceedings of JEP, Avignon, France.
- Audibert, N., Vincent, D., Aubergé, V., & Rosec, O. (2006). *Expressive speech synthesis: Evaluation of a voice quality centered coder on the different acoustic dimensions*. In proceedings of the Speech Prosody, Dresden, Germany.
- Averill, J. R. (1980). A constructivist view of emotion. In R. Plutchik & H. Kellerman (Eds.), *Emotion: Theory, Research and Experience* (Vol. 1, pp. 305-399). London: Academic Press.
- Bänziger, T., & Scherer, K. R. (2003). *Relations entre caractéristiques vocales perçues et émotions attribuées*. In proceedings of the Les Journées Prosodie.
- Banse, R., & Scherer, K. R. (1996). Acoustic Profiles in Vocal Emotion Expression. *Journal of personality and social psychology*, 70(3), 614-636.
- Bard, P. (1928). A diencephalic mechanism for the expression of rage with special reference to the sym- pathetic nervous system. *American Journal of Psychology*, 84, 490-515.
- Barkhuysen, P., Krahmer, E., & Swerts, M. (2007). *Cross-modal perception of emotional speech*. In proceedings of the ICPhS, Saarbruecken, Germany.
- Bassano, D., Maillochon, I., & Eme, P. (1998). Developmental changes and variability in early lexicon, a study of French children's naturalistic productions. *Journal of Child Language*, 25, 193-531.
- Batliner, A., Steidl, S., Hacker, C., Nöth, E., & Niemann, H. (2005). *Tales of tuning-prototyping for automatic classification of emotional user states*. In proceedings of the INTERSPEECH.
- Belotel-Grenié, A., & Grenié, M. (1997). Types de phonation et tons en chinois standard. *Cahiers de Linguistique – Asie Orientale*, 26 (2), 249-279.
- Benesty, J., Sondhi, M. M., & Huang, Y. (2008). *Springer handbook of speech processing*. Berlin: Springer Verlag.
- Berlin, B., & Kay, P. (1969). *Basic Color Terms: Their Universality and Evolution*. California: University of California Press.
- Bernicot, J. (1992). *Les actes de langage chez l'enfant*. Paris: PUF.
- Billières, M. (2002). La genèse du crible prosodique. Conséquences sur l'enseignement de la prononciation en langue étrangère. *Cahiers du Centre Interdisciplinaire des Sciences du Langage*, 16, 51-64.
- Billières, M. (2005). Codage phonologique et boucle articulatoire en mémoire de travail: un support pour la facilitation de l'accès à l'oral et à la lecture pour des publics débutants en français langue étrangère. *CORELA, Numéro spécial: Colloque AFLS*.
- Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glott International*, 5(9/10), 341-345.
- Bolinger, D. (1964). *Intonation as a universal*. In proceedings of the The 5th Congress of Phonetics, Cambridge.

- Bolinger, D. (1978). Intonation across languages. In J. H. Greenberg, C. A. Ferguson & E. A. Moravcsik (Eds.), *Universals of human language* (Vol. 2, pp. 471-524). Stanford: University Press.
- Bornstein, M., Kessen, W., & Weiskopf, S. (1976). The categories of hue in infancy. *Science*, 191(4223), 201-202.
- Brazill, T. J., Romney, A. K., & Batchelder, W. H. (1995). A Comparison of Methods for Collecting Judged Similarities among Items in a Semantic Domain. *Journal of Quantitative Anthropology*, 5, 359-374.
- Breckler, S. J. (1984). Empirical validation of affect, behavior, and cognition as distinct attitude components. *Journal of personality and social psychology*, 47, 1191-1205.
- Brunelle, M. (2003). *Coarticulation effects in Northern Vietnamese tones*. In proceedings of the 15th ICPHS, Barcelona.
- Caelen-Haumont, G. (1991). *Stratégies des locuteurs en réponse à des consignes de lecture d'un texte: Analyse des interactions entre modèles syntaxiques, sémantiques, pragmatiques et paramètres prosodiques*. (Thèse de doctorat d'état), Université de Provence, Aix-en-Provence.
- Caelen-Haumont, G., & Bel, B. (2000). Le caractère spontané dans la parole et le chant improvisés: de la structure intonative au mélisme. *Revue Parole*, 15(16), 251-302.
- Caelen-Haumont, G., & Keller, E. (1997). La prosodie, de la parole à la synthèse: l'apport de la sémantique et de la pragmatique. *Etudes de lettres*, 3, 103-130.
- Callamand, D. (1991). *Grammaire vivante du français : français langue étrangère*. Paris: Cle International.
- Campbell, D. T. (1950). The indirect assessment of social attitudes. *Psychological Bulletin*, 47, 15-38.
- Campbell, N. (2004). *Perception of affect in speech-towards an automatic processing of paralinguistic information in spoken conversation*. In proceedings of the INTERSPEECH.
- Campbell, N. (2005). Getting to the Heart of the Matter: Speech as the Expression of Affect; Rather than Just Text or Language. *Language Resources and Evaluation*, 39(1), 109-118. doi: 10.1007/s10579-005-2699-y
- Campbell, N., & Mokhtari, P. (2003). *Voice quality: The 4th prosodic dimension*. In proceedings of the The 15th international congress of phonetic sciences.
- Cannon, W. B. (1927). The James-Lange theory of emotions : A critical examination and an alternative theory. *American Journal of Psychology*, 39, 106-124.
- Cao, J. (2002). 汉语声调与语调的关系 (Relation entre le ton et l'intonation du chinois mandarin). *中国语文*(3), 195-202.
- Chang, Y. C. (1998). Les indices acoustiques et perceptifs des questions totales en Mandarin parle de Taiwan. *Cahiers de linguistique-Asie orientale*, 27(1), 51-78.
- Chao, Y. R. (1968). *A grammar of spoken Chinese*. California: University of California Press.
- Chen, F., Li, A., Wang, H., Wang, T., & Fang, Q. (2004). *Acoustic analysis of friendly speech*. In proceedings of the ICASSP.

- Cheng, X. (2013). *普通话态度语音韵律特征分析与感知实验 (Analyse prosodique et la perception des expressions attitudinales en chinois mandarin)*. (Mémoire de Master), Université normale de Nanjing.
- Chuang, C. K., Hiki, S., Sone, T., & Nimura, T. (1972). *The acoustical features and perceptual cues of the four tones of Standard Colloquial Chinese*. In proceedings of the Seventh International Congress on Acoustics, Adadémial Kiado, Budapest.
- Clément, J. (1999). *Structure des représentations prosodiques. Développement normal et pathologique du traitement de la prosodie*. (Thèse de doctorat), Université de Paris V.
- Cornelius, R. R. (1996). *The science of emotion: Research and tradition in the psychology of emotions*. Upper Saddle River (NJ): Prentice-Hall, Inc.
- Cornelius, R. R. (2000). *Theoretical approaches to emotion*. In proceedings of the SCA Tutorial and Research Workshop (ITRW) on Speech and Emotion.
- Couper-Kuhlen, E. (1986). *An Introduction to English Prosody*. London: Edward Arnold.
- Crary, M. A., & Tallman, V. L. (1993). Production of linguistic prosody by normal and speech disordered children. *Journal of Communication Disorders*, 26, 245-262.
- Cruttenden, A. (1986). *Intonation*. Cambridge Cambridge University Press.
- Cruttenden, A. (1997). *Intonation*. Cambridge: Cambridge University Press.
- Crystal, D. (1969). *Prosodic Systems and Intonation in English*. Cambridge: Cambridge University Press.
- Crystal, D. (1975). *The English tone of voice*. London: Edward Arnold.
- Crystal, D. (1986). Prosodic development. In P. Fletcher & M. Garman (Eds.), *Language acquisition: studies in first language development (2nd edition)* (pp. 174-197). Cambriadge: Cambridge University Press.
- d'Alessandro, C. (2006). Voice source parameters and prosodic analysis. In S. Sudhoff, D. Lenertová, R. Meyer, S. Pappert, P. Augurzky, I. Mleinek, N. Richter & J. Schließer (Eds.), *Methods in Empirical Prosody Research* (pp. 63-87). Berlin-New York: Walter de Gruyter.
- d'Alessandro, C., Rilliard, A., & Le Beux, S. (2011). Chironomic stylization of intonationa. *The Journal of the Acoustical Society of America*, 129(3), 1594-1604.
- Danes, F. (1994). Involvement with language and in language. *Journal of pragmatics*, 22(3-4), 251-264.
- Daneš, F. (1960). Sentence intonation from a functional point of view. *Word*, 16, 34-54.
- de Boysson-Bardies, B. (1996). *Comment la parole vient aux enfants*. Paris: Odile Jacob.
- de Meo, A. (2012). *How credible is a non-native speaker? Prosody and surroundings*. In proceedings of the Workshop of Methodological Perspective on Second Language Prosody.
- de Moraes, J. A., & Rilliard, A. (2014). Illocution, Attitudes and Prosody. In T. Raso (Ed.), *Spoken Corpora and Linguistic Studies*. Amsterdam: John Benjamins Publisher.

- de Moraes, J. A., Rilliard, A., Alberto, B., & Shochi, T. (2010). *Production and perception of attitudinal meaning in Brazilian Portuguese*. In proceedings of the Speech Prosody, Chicago, USA.
- Delattre, P. (1966). Les dix intonations de base du français. *The French Review*, 40(1), 1-14.
- Deller, J. R., Proakis, J. G., & Hansen, J. H. (2000). *Discrete-time processing of speech signals*. In proceedings of the IEEE, New York, USA.
- Deng, L., & Dang, J. (2007). Speech Analysis: The Production –Perception Perspective. In H. Li & C. Lee (Eds.), *Advances in Chinese Spoken Language Processing* (pp. 3-32). New Jersey: Word Scientific Publisher.
- Detey, S. (2005). Utiliser l'écrit au service de l'oral. *Le français dans le monde*, 342, 38-40.
- Devillers, L., & Vidrascu, L. (2006). *Real-life emotions detection with lexical and paralinguistic cues on human-human call center dialogs*. In proceedings of the INTERSPEECH.
- Di Cristo, A. (1975). *Soixante et dix ans de recherches en prosodie (Vol. 1)*. Aix-en-Provence: Publications de l'Université de Provence.
- Di Cristo, A. (2013). *La prosodie de la parole*. Paris: De Boeck-Solal.
- Diaferia, M. L. (2002). *Les Attitudes de l'Anglais : Premiers Indices Prosodiques*. (Mémoire de master), National Polytechnique Institut of Grenoble, France.
- Dore, J. (1975). Holophrases, speech acts, and language universals. *Journal of Child Language*, 2, 21-40.
- Douglas-Cowie, E., Campbell, N., Cowie, R., & Roach, P. (2003). Emotional speech: Towards a new generation of databases. *Speech communication*, 40(1), 33-60.
- Dreher, J. J., & Lee, P. C. E. (1968). Instrumental investigation of single and paired Mandarin tonemes. *Monumenta serica*, 343-373.
- Duanmu, S. (2002). *The Phonology of Standard Chinese*. Oxford: Oxford University Press.
- Dumas, G. (1948). *La Vie affective, Physiologie - Psychologie - Socialisation*. Paris: P.U.F.
- Durst, U. (2003). The Natural Semantic Metalanguage Approach to linguistic meaning. *Theoretical Linguistics*, 3(29), 157-200.
- Ekman, P. (1989). The argument and evidence about the universals in facial expressions of emotion. *Handbook of social psychophysiology*, 143-164.
- Ekman, P. (1999). Facial expressions. In T. Dalgleish & M. Power (Eds.), *Handbook of Cognition and Emotion* (pp. 301-320). Sussex, U.K.: John Wiley & Sons, Ltd.
- Ekman, P., & Friesen, W. V. (1978). *Facial Action Coding System*. Palo Alto, CA: Consulting Psychologists Press.
- Ellis, R. (2000). *Second Language Acquisition*. Oxford: Oxford University Press.
- Fónagy, I. (1956). Über die Eigenart des Sprachlichen Zeichens. *Lingua*, 6, 67-88.
- Fónagy, I. (1969). Métaphore d'intonation et changement d'intonation. *Bulletin de la Société de Linguistique de Paris*, 64, 22-42.
- Fónagy, I. (1971). Double coding in speech. *Semiotica*, 3, 189-222.

- Fónagy, I. (1983). *La vive voix: essai de psychophonétique*. Paris: Payot.
- Fónagy, I., Bérard, E., & Fónagy, J. (1983). Clichés mélodiques. *Folia linguistica*, 17(1-4), 153-186.
- Fazio, R. H., Eiser, J. R., & Shook, N. J. (2004). Attitude formation through exploration: Valence asymmetries. *Journal of personality and social psychology*, 87, 293-311.
- Fernald, A., & Mazzie, C. (1991). Prosody and focus in speech to infants and adults. *Developmental Psychology*, 27(209-213).
- Fu, R. (2006). Politiques et stratégies linguistiques dans l'enseignement supérieur des langues étrangères en Chine nouvelle. *Synergie Chine*, 1, 27-39.
- Fujisaki, H., & Hirose, K. (1993). Analysis and perception of intonation expressing paralinguistic information in spoken Japanese. *ESCA Workshop on Prosody*.
- Fuson, W. M. (1942). Attitude : A note on the concept and its research consequences. *American Sociological Review*, 7, 856-857.
- Gårding, E. (1987). Speech act and tonal pattern in Standard Chinese: constancy and variation. *Phonetica*, 44(1), 13-29.
- Gandour, J. (1984). Tone dissimilarity judgments by Chinese listeners. *Journal of Chinese Linguistics*, 12, 235-261.
- Gelb, I. J. (1973). *Pour une théorie de l'écriture*. Paris: Flammarion.
- Gergen, K. J. (1985). The social constructionist movement in modern psychology. *American Psychologist*, 40, 266-275.
- Gobl, C., & Ní Chasaide, A. (2000). *Testing affective correlates of voice quality through analysis and resynthesis*. In proceedings of the ISCA Workshop on Speech and Emotions, Newcastle, Irlande du Nord.
- Gobl, C., & Ní Chasaide, A. (2003). The role of voice quality in communicating emotion, mood and attitude. *Speech communication*, 40(1), 189-212.
- Grépillat, T. (1996). *Perçoit-on, par l'intonation, l'attitude d'un locuteur avant la fin de l'énoncé?* (Mémoire de master), Université Stendhal Grenoble III.
- Grandjean, D., & Scherer, K. R. (2009). Théorie de l'évaluation cognitive et dynamique des processus émotionnels. In D. Sander & K. R. Scherer (Eds.), *Traité de psychologie des émotions* (pp. 42-76). Paris: Dunod.
- Granström, B., & House, D. (2005). Audiovisual representation of prosody in expressive. *Speech communication*, 46(3), 473-484.
- Grosz, B., & Hirschberg, J. (1992). *Some intonational characteristics of discourse structure*. In proceedings of the Second International Conference on Spoken Language Processing (ICSLP'92), Banff, Canada.
- Gu, W., Zhang, T., & Fujisaki, H. (2011). *Prosodic Analysis and Perception of Mandarin Utterances Conveying Attitudes*. In proceedings of the INTERSPEECH.
- Gussenhoven, C. (2004). *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press.
- Halliday, M. A. K. (1970). *A Course in Spoken English: Intonation*. Oxford: Oxford University Press.

- Halliday, M. A. K. (1975). *Learning how to mean – exploration in the development of language*. Londres: Arnold.
- Han, M. S., & Kim, K. O. (1974). Phonetic variation of Vietnamese tones in disyllabic utterances. *Journal of Phonetics*, 2, 223-232.
- Hermes, D. J. (1998). Measuring the perceptual similarity of pitch contours. *Journal of Speech, Language, and Hearing Research*, 41(1), 73-82.
- Hirschfeld, U., & Trouvain, J. (2007). Teaching prosody in German as a foreign language. *Trends in Linguistics Studies and Monographs*, 186, 171-187.
- Hirst, D., & Di Cristo, A. (1998). *Intonation systems: a survey of twenty languages*. Cambridge: Cambridge University Press.
- Ho, A. T. (1976). Mandarin tones in relation to sentence intonation and grammatical structure. *Journal of Chinese Linguistics*, 4(1).
- Ho, A. T. (1977). Intonation variation in a Mandarin sentence for three expressions: interrogative, exclamatory and declarative. *Phonetica*, 34(6), 446-457.
- Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal*, 50(3), 346-363.
- Howie, J. M. (1970). *The vowels and tones of Mandarin Chinese: Acoustical measurements and experiments*. (Thèse de doctorat), Indiana University.
- Howie, J. M. (1976). *Acoustical studies of Mandarin vowels and tones*. Cambridge: Cambridge University Press.
- Husson, F., Josse, J., & Pages, J. (2010). Principal component methods-hierarchical clustering-partitional clustering: why would we need to choose for visualizing data? *Rapport technique- Agrocampus*. Département des mathématiques appliquées. <http://www.agrocampus-ouest.fr/math/>.
- James, W. (1884). What is an emotion? *Mind*, 9, 188-205.
- James, W. (1894). Discussion: The physical basis of emotion. *Psychological Review*, 1(5), 516.
- Johnstone, T., & Schere, K. R. (1999). *The effects of emotions on voice quality*. In proceedings of the The 14th International Congress of Phonetic Science, San Francisco, Etats-Unis.
- Kay, P., & Maffi, L. (1999). Color Appearance and the Emergence and Evolution of Basic Color Lexicons. *American Anthropologist*, 101(4), 743-760.
- Kehrein, R. (2002). *The prosody of authentic emotions*. In proceedings of the Speech Prosody.
- Knapp, A. (2006). Phonologie et prosodie dans l'acquisition de l'allemand-L2 par des apprenants francophones. *Zeitschrift für Interkulturellen Fremdsprachenunterricht [Online]*, 9(2), 29 pp.
- Konopczynski, G. (1990). *Le langage émergent I : caractéristiques rythmiques, et II : aspects vocaux et mélodiques*. Hambourg: Helmut Buske Verlag.
- Konopczynski, G. (2000). The development interactive intonology model : applications to French. *Parole*, 7/8, 177-201.

- Kramer, E. (1964). Elimination of verbal cues in judgments of emotion from voice. *Journal of Abnormal and Social Psychology*, 68, 390-396.
- Krech, D., & Crutchfield, R. S. (1948). *Theory and problems of social psychology*. New York: McGraw-Hill.
- Kumbasar, E., Romney, A. K., & Batchelder, W. H. (1994). Systematic bias in social perception. *American Journal of Sociology*, 100(2), 477-505.
- Léon, P. (1971). Essai de phonostylistique. *Studia Phonetica*, 4, Montréal: Marcel Didier, Inc, 185 p.
- Léon, P. (1993). *Précis de phonostylistique: parole et expressivité*. Paris: NATHAN.
- Léon, P. R. (2000). *Phonétisme et prononciation du français*. Paris: Nathan.
- Lacheret-Dujour, A., & Beaugendre, F. (1999). *La prosodie du français*. Paris: Editions CNRS.
- Ladd, D. R. (1984). Declination: a review and some hypotheses. *Phonology*, 1, 53-74.
- Lange, C. G. (1885/1967). The emotions. In C. G. L. a. W. James (Ed.). New York: Hafner.
- Laver, J. (1980). *The Phonetic Description of Voice Quality*. Cambridge: Cambridge University Press.
- Lazarus, R. S. (1966). *Psychological Stress and Coping Process*. New York: McGraw Hill.
- Li, A. (2005). Acoustic analysis on friendly speech. *Chinese Language*, 5, 004.
- Li, A., Fang, Q., & Dang, J. (2011). *Emotional intonation in a tone language: Experimental evidence from Chinese*. In proceedings of the ICPhS XVII, Hong Kong.
- Li, A., & Wang, H. (2004). *Friendly Speech Analysis and Perception in Standard Chinese*. In proceedings of the INTERSPEECH.
- Li, C. N., & Tompson, S. A. (1981). *Mandarin Chinese: a functional reference grammar*: University of California Press.
- Li, Y-H. A. (1990). *Order and constituency in Mandarin Chinese*. Dordrecht: Kluwer Academic Publishers.
- Li, Z. (1990). Méthodes d'enseignement du français, seconde langue étrangère en Chine. In A. Obadia (Ed.), *Premier colloque international sur l'enseignement du français en Chine: communications choisies* (pp. 145-152). Canada: Les Presses de l'Université Simon Fraser.
- Lieberman, P., & Michaels, S. B. (1962). Some aspects of fundamental frequency and envelope amplitude as related to the emotional content of speech. *Journal of the Acoustical Society of America*, 34, 922-927.
- Lin, H. Y., & Fon, J. (2012). *Prosodic and Acoustic Features of Emotional Speech in Taiwan Mandarin*. In proceedings of the Speech Prosody.
- Lin, M. (1988). 普通话声调的声学特征和知觉征兆 (Caractéristiques acoustiques et indices perceptifs des tons en chinois mandarin). *中国语文*, 3, 182-193.
- Lin, M. (2002). 普通话语句的韵律结构和基频 (F0) 高低线构建 (Structure prosodique du chinois mandarin et les lignes haut et bas du contour F0). *当代语言学*, 4(4), 254-265.

- Lin, M., & Yan, J. (1992). Tonal coarticulation patterns in quadrisyllabic word and phrase of mandarin. *Acta Acustica*, 6(008).
- Loveday, L. (1981). Pitch, politeness and sexual role: An exploratory investigation into the pitch correlates of English and Japanese politeness formulae. *Language and Speech*, 24(1), 71-89.
- Lu, M., & Yang, W. (2012). Multivariate Logistic Regression Analysis of Complex Survey Data with Application to BRFSS Data. *Journal of Data Science*, 10(2), 157-173.
- Lucy, J. A. (1997). The linguistics of "color". In C. L. Hardin & L. Maff (Eds.), *Color categories in thought and language* (pp. 320-346). Cambridge: Cambridge University Press.
- Mac, D. K. (2012). *Génération de parole expressive dans le cas des langues à tons*. Thèse de doctorat, Université de Grenoble.
- Mac, D. K., Aubergé, V., Castelli, E., & Rilliard, A. (2012). *Local vs. global prosodic cues: effect of tones on attitudinal prosody in cross-perception of Vietnamese by French*. In proceedings of the Speech Prosody.
- Mac, D. K., Aubergé, V., Rilliard, A., & Castelli, E. (2010). *How prosodic attitudes can be recognized and confused: Vietnamese multimodal social affects*. In proceedings of the SLTU, Penang, Malaysia.
- Malrieu, P. (1952). Les émotions et la personnalité de l'enfant. *Études de Psychologie et de Philosophie*, XII. Paris Librairie philosophique.
- Martin, F. (1990). Note sur l'histoire de la série des quatre tons. *Extrême orient Extrême occident*, 12(12), 67-78.
- Martin, P. (2009). *Intonation du français*. Paris: Armand Colin.
- Martinet, A. (1969). *Langue et fonction: une théorie fonctionnelle du langage*. Paris: Denoël/Gonthier.
- Martinet, A. (1970). *Éléments de linguistique générale*. Paris: Armand Colin.
- Martins-Baltar, M. (1977). *De l'énoncé à l'énonciation: une approche des fonctions intonatives*. Paris: Didier.
- Massaro, D. W., Cohen, M. M., & Tseng, C. C. (1985). The evaluation and integration of pitch height and pitch contour in lexical tone perception in Mandarin Chinese. *Journal of Chinese Linguistics*, 13, 267-289.
- Mehler, J., Bertoncini, J., Barrière, M., & Jassik-Gerschenfeld, D. (1978). Infant recognition of mother's voice. *Perception*, 7(5), 491-497.
- Mejvaldová, J. (2000). *Expressions prosodiques de certaines attitudes en thème et en français: étude comparative*. Paris: Université Paris 7 – Denis Diderot.
- Michard, C. (2002). *Sexe en linguistique: Sémantique ou Zoologie?* Paris: L'Harmattan.
- Michaud, A. (2008). *Tones and intonation: some current challenges*. In proceedings of the The 8th international seminar on speech production (ISSP'08), Strasbourg.
- Moore, C. B., & Jongman, A. (1997). Speaker normalization in the perception of Mandarin Chinese tones. *The Journal of the Acoustical Society of America*, 102(3), 1864-1877.



- Moore, C. C., Romney, A. K., & Hsia, T. L. (2002). Cultural, gender, and individual differences in perceptual and semantic structures of basic colors in Chinese and English. *Journal of Cognition and Culture*, 2(1), 1-28.
- Moore, C. C., Romney, A. K., Hsia, T. L., & Rusch, C. D. (1999). The Universality of the Semantic Structure of Emotion Terms: Methods for the Study of Inter - and Intra - Cultural Variability. *American Anthropologist*, 101(3), 529-546.
- Morlec, Y. (1997). *Génération multiparamétrique de la prosodie du français par apprentissage automatique*. (Thèse de doctorat), Institut National Polytechnique de Grenoble.
- Morlec, Y., Bailly, G., & Aubergé, V. (2001). Generating prosodic attitudes in French: data, model and evaluation. *Speech communication*, 33(4), 357-371.
- Moutinho, L. C., Coimbra, R. L., Rilliard, A., & Romano, A. (2011). Mesure de la variation prosodique diatopique en portugais européen. *Estudios de Fonética Experimental*, 33-55.
- Mozziconacci, S. (2001). *Emotion and attitude conveyed in speech by means of prosody*. In proceedings of the The 2nd Workshop on Attitude, Personality and Emotions in User-Adapted Interaction, Sonthofen, Germany.
- Nordenhake, M., & Svantesson, J. O. (1983). Duration of Standard Chinese word tones in different sentence environments. *Lund University Working papers in Linguistics*, 25, 105-111.
- Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2005). Understanding and using the implicit association test: II. Method variables and construct validity. *Personality and Social Psychology Bulletin*, 31, 166-180.
- Nugier, A. (2009). Histoire et grands courants de recherche sur les émotions. *Revue électronique de Psychologie Sociale*(4), 8-14.
- O'Connor, J. D., & Arnold, G. F. (1978). *Intonation of Colloquial English*. London: Longman.
- Ohala, J. J. (1983). Cross-Language Use of Pitch: An Ethological View. *Phonetica (International Journal of Phonetic Science)*, 40(1), 1983
- Ohala, J. J. (1994). The frequency codes underlies the sound symbolic use of voice pitch. In L. Hinton, J. Nichols, & J. J. Ohala (Eds.), *Sound symbolism* (pp. 325-347). Cambridge: Cambridge University Press.
- Ohala, J. J. (1996). *Ethological theory and the expression of emotion in the voice*. In proceedings of the Fourth International Conference on Spoken Language.
- Ohara, Y. (1992). *Gender dependent pitch levels: A comparative study in Japanese and English*. In proceedings of the Second Berkeley women and language conference.
- Pao, T. L., Chen, Y. T., Yeh, J. H., & Lu, J. J. (2005). Detecting Emotions in Mandarin Speech. *Computational Linguistics and Chinese Language Processing*, 10(3), 347-362.
- Passy, P. (1890/1979). *Etudes sur les changements phonétiques*. Paris: Didrot.
- Pavlenko, A. (2007). *Emotions and multilingualism*. Cambridge: Cambridge University Press.
- Pernet-Liu, A. (2013). Les sujets de mémoires d'étudiants chinois: choix personnel et culture universitaire. *Synergie Chine*, 8, 177-184.

- Pettorino, M., De Meo, A., & Vitale, M. (2012). *Transplanting Credibility into a Foreign Voice. An Experiment on Synthesized L2 Italian*. In proceedings of the Speech and Corpora. Proceedings of the 7th GSCP International Conference, Firenze.
- Pettorino, M., De Meo, A., Pellegrino, E., Salvati, L., & Vitale, M. (2011). Accento straniero e credibilità del messaggio: un'analisi acustico-percettiva. In B. Gili Fivela, A. Stella, L. Garrapa & M. Grimaldi (Eds.), *Contesto comunicativo e variabilità nella produzione e percezione della lingua, Atti del 7° Convegno Nazionale dell'Associazione Italiana di Scienze della Voce (AISV 2011)*. Roma: Bulzoni editore.
- Piske, T., MacKay, I., & Flege, J. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics*, 29, 191-215.
- Pu, Z., Lu, J., & Xu, X. (2006). Survol historique des manuels de français en Chine. *Synergie Chine*, 1, 72-79.
- Raphael, L. J., Borden, G. J., & Harris, K. S. (2007). *Speech science primer: Physiology, acoustics, and perception of speech*. Lippincott Williams & Wilkins.
- Rilliard, A., Erickson, D., de Moraes, J. A., & Takaaki, S. (2014). Cross-cultural perception of some Japanese politeness and impoliteness expressions. In F. Baider & G. Cislariu (Eds.), *Linguistic Approaches to Emotions in Context* (pp. 251-276).
- Rilliard, A., & Lai, J. P. (2008). *Outils pour le calcul et la comparaison prosodique dans le cadre du projet AMPER-L'exemple des variétés occitane et sarde*. In proceedings of Symposium International "La variation diatopique de l'intonation dans le domaine roumain et roman" (pp. 217-229). Iași, Roumanie: Université Alexandre-Jean-Cuza de Iași.
- Rilliard, A., Martin, J. C., Aubergé, V., & Shochi, T. (2008). *Perception of French Audio-Visual Prosodic Attitudes*. In proceedings of the Speech Prosody, Campinas, Brasil.
- Rilliard, A., Shochi, T., Martin, J. C., Erickson, D., & Auberge, V. (2009). Multimodal Indices to Japanese and French Prosodically Expressed Social Affects. *Language and Speech*, 52(2-3), 223-243.
- Romano, A., & Miotti, R. (2008). *Distancias prosódicas entre variedades románicas*. In proceedings of Symposium International "La variation diatopique de l'intonation dans le domaine roumain et roman" (pp. 231-249). Iași, Roumanie: Université Alexandre-Jean-Cuza de Iași.
- Romney, A. K., Batchelder, W. H., & Brazil, T. (1995). Scaling semantic domains. In R. D. Luce, M. D'Zmura, D. Hoffman, G. J. Iverson & A. K. Romney (Eds.), *Geometric representations of perceptual phenomena* (pp. 267-294). Hillsdale, NJ: Lawrence Erlbaum.
- Romney, A. K., Boyd, J. P., Moore, C. C., Batchelder, W. H., & Brazill, T. J. (1996). Culture as Shared Cognitive Representations. *Proceedings of the National Academy of Sciences*, 93(10), 4699-4705.
- Romney, A. K., & Moore, C. C. (1998). Toward a theory of culture as shared cognitive structures. *Ethos*, 26(3), 314-337.
- Romney, A. K., Moore, C. C., Batchelder, W. H., & Hsia, T. L. (2000). Statistical methods for characterizing similarities and differences between semantic structures. *The National Academy of Sciences*, 97(1), 518-523.

- Romney, A. K., Moore, C. C., & Rusch, C. D. (1997). Cultural universals: Measuring the semantic structure of emotion terms in English and Japanese. *The National Academy of Sciences*, 94(10), 5489-5494.
- Roseberry-Mckibbin, C., & Brice, A. (1999). The perception of vocal cues of emotion by Spanish-speaking limited English proficient children. *Journal of Children's Communication Development*, 20(2), 19-24.
- Rosenberg, M. J., & Hovland, C. I. (1960). Cognitive, affective, and behavioral components of attitudes. In C. I. Hovland & M. J. Rosenberg (Eds.), *Attitude organization and change: An analysis of consistency among attitude components* (pp. 1-14). New Haven, Conn: Yale University Press.
- Ross, E. D. (2000). Affective Prosody and the Aprosodias. In M. Mesulam (Ed.), *Principles of Behavioral and Cognitive Neurology* (pp. 316-331). New York: Oxford University Press.
- Ross, E. D., Edmondson, J. A., & Seibert, G. B. (1986). The effect of affect on various acoustic measures of prosody in tone and non-tone languages: A comparison based on computer analysis of voice. *Journal of Phonetics*, 14, 283-302.
- Rossi, M. (1999). *L'intonation: le système du français : description et modalisation*: Editions OPHRYS.
- Sadanobu, T. (2004). A natural history of Japanese pressed voice. *Journal of the Phonetic Society of Japan*, 8(1), 29-44.
- Saunders, B. (1995). Disinterring Basic Color Terms: a study in the mystique of cognitivism. *History of the Human Sciences*, 8 (7), 19-38.
- Scherer, K. R. (1984). On the nature and function of emotion: A component process approach. In K. R. Scherer & P. Ekman (Eds.), *Approaches to emotion* (pp. 293-317). Hillsdale: Erlbaum.
- Scherer, K. R. (1986). Vocal Affect Expression: A Review and a Model for Future Research. *Psychological Bulletin*, 99 (2), 143-165.
- Scherer, K. R. (2001). Appraisal considered as a process of multi-level sequential checking. In K. Scherer, A. Shorr & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, Methods, Research* (pp. 92-12): Oxford University Press.
- Scherer, K. R., Bänziger, T., & Grandjean, D. (2003). L'étude de l'expression vocale des émotions: mise en évidence de la dynamique des processus affectifs. In J. M. Colletta, & Tcherkassof, A. (Ed.), *Les Emotions: Cognition, langage et développements* (Vol. 247, pp. 39-58). Bruxelles: Editions Mardaga.
- Scherer, K. R., Banse, R., & Wallbott, H. G. (2001). Emotion inferences from vocal expression correlate across languages and cultures. *Journal of Cross-cultural psychology*, 32(1), 76-92.
- Scherer, K. R., & Ellgring, H. (2007). Multimodal expression of emotion: affect programs or componential appraisal patterns? *Emotion*, 7(1), 158-171.
- Scherer, K. R., Ladd, D. R., & Silverman, K. E. (1984). Vocal cues to speaker affect: Testing two models. *The Journal of the Acoustical Society of America*, 76(5), 1346-1356.
- Schubiger, M. (1958). *English Intonation: Its Form and Function*. Tübingen, Germany: Niemeyer.

- Seighworth, G. J., & Gregg, M. (2010). An Inventory Shimmers. In M. Gregg & G. J. Seigworth (Eds.), *The affect theory reader* (pp. 1-25). Durham & London: Duke University Press.
- Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics*, 10, 219-213.
- Selinker, L. (1992). *Rediscovering Interlanguage*. London: Longman.
- Seyfert, R. (2012). Beyond personal feelings and collective emotions: toward a theory of social affect. *Theory, Culture & Society*, 29(6), 27-46.
- Shen, J. (1985). 北京话声调的音韵和语调 (Prosodie et intonation des tons au dialect de Pékin). *北京语音实验*, 73-130.
- Shen, X. (1990). Tonal coarticulation in Mandarin. *Journal of Phonetics*, 18, 281-285.
- Shi, P. (1980). 四种句子的语调变化 (Intonation déclarative, interrogative, impérative et exclamative du chinois mandarin). *语言教学与研究*, 2, 71-81.
- Shochi, T. (2008). *Prosodie des affects socioculturels en japonais, français et anglais : à la recherche des vrais et faux-amis pour le parcours de l'apprenant*. Université de Stendhal - Grenoble III.
- Shochi, T., Aubergé, V., & Rilliard, A. (2005). Because Attitudes Are Social Affects, They Can Be False Friends. *Affective Computing and Intelligent Interaction*, 482-489.
- Shochi, T., Aubergé, V., & Rilliard, A. (2006). *How prosodic attitudes can be false friends: Japanese vs. French social affects*. In proceedings of the Speech Prosody, Dresden.
- Shochi, T., Aubergé, V., & Rilliard, A. (2007). *Cross-Listening of Japanese, English and French social affect : about universals, false friends and unknown attitudes*. In proceedings of the ICPhS 2007, Saarbrücken, Germany.
- Shochi, T., Gagné, G., Rilliard, A., Erickson, D., & Auberge, V. (2010). *Learning effect of prosodic social affects for Japanese learners of French language*. In proceedings of the Speech Prosody.
- Shochi, T., Rilliard, A., Aubergé, V., & Erickson, D. (2009). Intercultural Perception of English, French and Japanese Social Affective Prosody. In S. Hancil (Ed.), *The role of prosody in Affective Speech* (97) (pp. 31-59).
- Soni Madhulatha, T. (2012). An Overview on Clustering Method. *Journal of Engineering*, 2(4), 719-725.
- Sun, C. T., & Givon, T. (1985). On the so-called SOV word order in Mandarin Chinese: A quantified text study and its implications. *Language*, 61(2), 329-351.
- Swerts, M., & Krahmer, E. (2005). Audiovisual prosody and feeling of knowing. *Journal of Memory and Language*, 53(1), 81-94.
- t'Hart, J., & Cohen, A. (1973). Intonation by rule: a perceptual quest. *Journal of Phonetics*, 1, 309-327.
- Tao, J., & Kang, Y. (2006). 情感语音韵律自动生成方法的研究 (Etude de la génération automatique de la prosodie des émotions en chinois mandarin). *中国声学学会 2006 年全国声学学术会议论文集*.

- Tcherkassof, A. (2009). *Les émotions et leurs expressions*. Grenoble: Presse Universitaire de Grenoble.
- Trouvain, J., & Gut, U. (2007). *Non-native prosody: phonetic description and teaching practice*. Berlin: De Gruyter Mouton.
- Trudgill, P. (1983). *Sociolinguistics: An introduction to language and society*. London: Penguin Books.
- Tseng, C. C. (1981). *An acoustic phonetic study on tones in Mandarin Chinese*. (Thèse de doctorat), Brown University.
- Vaissière, J. (1997). Ivan Fonagy et la notion prosodique. In I. Fonagy, & Perrot, J. (Ed.), *Polyphonie pour Ivan Fonagy* (pp. 479-488). Paris: Editions L'Harmattan.
- Vaissière, J. (2006). *La phonétique*. Paris: Presses universitaires de France.
- Vallerand, R. J., & Lafrenaye, Y. (2006). Les attitudes. In R. J. Vallerand (Ed.), *Les fondements de la psychologie sociale* (pp. 235-291). Montréal: Gaétan Morin éditeur.
- Van Bezooijen, R., Otto, S., & Heenan, T. (1983). Recognition of vocal expressions of emotion. *Journal of Cross-cultural psychology*, 14, 387-406.
- Wallon, H. (1934/1970). *Les origines du caractère chez l'enfant* (4e ed. ed.). Paris: P. U. F.
- Wang, A., Chen, M., & Lu, S. (2002). 汉语陈述句音高下倾初探 (*Déclinaison des phrases déclaratives en chinois mandarin*). In proceedings of the 声调和语调研讨会.
- Wang, B., Yang, Y., & Lu, S. (2001). 汉语韵律层级边界结构的声学相关物 (*Structure hiérarchique de la prosodie du chinois mandarin et des caractéristiques acoustiques corrélées*). 新世纪的现代语音学——第五届现代语音学学术会议文集.
- Wang, H., & Li, A. (2003). 普通话情绪语音库的建立及听辨实验 (*Construction et la perception d'un corpus audio des émotions en chinois mandarin*). In proceedings of the 第六届全国现代语音学学术会议.
- Wang, M., & Lin, M. (2003). 普通话自然话语中的下倾 (*Déclinaison de la parole naturelle en chinois mandarin*). In proceedings of the 第六届全国现代语音学学术会议.
- Wichmann, A. (2000). *The attitudinal effects of prosody, and how they relate to emotion*. In proceedings of ISCA Tutorial and Research Workshop (ITRW) on Speech and Emotion: Developing a Conceptual Frame-work (pp. 143-147). Newcastle, Irelande.
- Wierzbicka, A. (1992). Talking about emotions: Semantics, culture, and cognition. *Cognition & Emotion*, 6(3-4), 285-319.
- Wierzbicka, A. (1996). *Semantics: Primes and Universals*. Oxford: Oxford University Press.
- Wierzbicka, A. (1999). *Emotions Across Languages and Cultures: Diversity and universals*. Cambridge: Cambridge University Press.
- Wierzbicka, A. (2005). There Are No Color Universals But There Are Universals of Visual Semantics. *Anthropological linguistics*, 47(2), 217-244.

- Wierzbicka, A. (2006). The Semantics of Colour: A New Paradigm. In N. J. Pitchford & C. P. Biggam (Eds.), *Progress in Colour Studies: Volume I. Language and Culture* (pp. 1–24). Philadelphia: John Benjamins Publishing Company.
- Wu, Z. (1982). Rules of intonation in Standard Chinese. *Papers for the working Group on Intonation ICL*, 13, 95-103.
- Wu, Z. (2000). *From Traditional Chinese Phonology to Modern Speech Processing : Realization of Tone and Intonation in Standard Chinese*. In proceedings of the ICSLP 2000.
- Xu, Y. (1999). Effects of tone and focus on the formation and alignment of f0 contours. *Journal of Phonetics*, 27(1), 55-105.
- Yang-Drocourt, Z. (2007). *Parlons chinois*. Paris: L'Harmattan.
- Yuan, J. (2004). *Perception of Mandarin intonation*. In proceedings of the International Symposium on Chinese Spoken Language Processing.
- Yuan, J., Shen, L., & Chen, F. (2002). *The acoustic realization of anger, fear, joy and sadness in Chinese*. In proceedings of the Seventh International Conference on Spoken Language Processing.
- Yuan, J. H. (2001). *Hanyu fangyan gaiyao [An introduction to Chinese Dialects]*. Beijing: Yuwen Press.
- Zanna, M. P., Haddock, G., & Esses, V. M. (1990). *The determinants of prejudice*. In proceedings of the The annual meeting of the Society for Experimental Social Psychology, Buffalo, NY.
- Zanna, M. P., & Rempel, J. K. (1988). Attitudes: A new look at an old concept. In D. Bar-Tal & A. Kruglanski (Eds.), *The social psychology of knowledge* (pp. 315-334). New York: Cambridge University Press.
- Zhang, F. (1990). Etude sur l'enseignement du français langue étrangère en Chine. In A. Obadia (Ed.), *Premier colloque international sur l'enseignement du français en Chine: communications choisies* (pp. 19-33). Canada: Les Presses de l'Université Simon Fraser.
- Zhang, S., Ching, P. C., & Kong, F. (2006). *Acoustic analysis of emotional speech in Mandarin Chinese*. In proceedings of the nternational Symposium on Chinese Spoken Language Processing.

# Annexe 1: Les énoncés du corpus d'affects sociaux chinois

Le corpus d'affects sociaux chinois contient au total 152 énoncés, construits selon les variations prosodiques suivantes : la longueur des énoncés (en syllabes), le placement des tons et la structure syntaxique des énoncés. La longueur des énoncés varie entre une et neuf syllabes. Les tons sont représentés par des chiffres : 1 fait référence à un ton plat, 2 à un ton montant, 3 à un ton descendant-remontant, 4 à un ton descendant et 0 à un ton neutre. La structure syntaxique varie sur la dimension paradigmatic et syntagmatic : l'unité lexicale (mot et groupe de mots), phrase canonique comprenant un sujet, un prédicat avec ou sans complément d'objet.

## Une syllabe : 8 énoncés

N°	Ton	Phrase en chinois	Transcription pinyin	Traduction française	Commentaire
1	1	书	shu1	livre	nom
2	2	门	men2	porte	nom
3	3	雨	yu3	pluie	nom
4	4	六	liu4	6	chiffre
5	1	书	shu1	livre	nom
6	2	熟	shu2	cuit	adjectif
7	3	数	shu3	compter	verbe
8	4	树	shu4	arbre	nom

## Deux syllabes : 20 énoncés

N°	Ton	Phrase en chinois	Transcription pinyin	Traduction française	Commentaire
1	1-1	声音	sheng1yin1	voix	nom
2	1-2	春节	chun1jie2	fête du printemps	nom
3	1-3	歌手	ge1shou3	chanteur	nom
4	1-4	三万	san1wan4	30 000	chiffre
5	1-0	他们	ta1men0	eux	pronom
6	2-1	明天	ming2tian1	demain	adverbe
7	2-2	诚实	cheng2shi2	honnête	Adjectif
8	2-3	滑雪	hua2xue3	ski	nom
9	2-4	营业	ying2ye4	affaires	nom
10	2-0	鼻子	bi2zi0	nez	nom
11	3-1	纸箱	zhi3xiang1	carton	nom

12	3-2	草莓	cao3mei2	fraise	nom
13	3-3	手表	shou3biao3	montre	nom
14	3-4	卡片	ka3pian4	carte	nom
15	3-0	我们	wo3men0	nous	pronom
16	4-1	六千	liu4qian1	6 000	chiffre
17	4-2	放学	fang4xue2	l'école est finie	verbe
18	4-3	电脑	dian4nao3	ordinateur	nom
19	4-4	世界	shi4jie4	monde	nom
20	4-0	钥匙	yao4shi0	clé (à la porte)	nom

### Trois syllabes : 62 énoncés

N°	Ton	Phrase en chinois	Transcription pinyin	Traduction française	Commentaire
1	2-1-0	抬桌子	tai2zhuo1zi0	lever une table	GV
2	2-2-0	胡萝卜	hu2luo2bo0	carotte	nom
3	2-3-0	量尺寸	liang2chi3cun0	mesurer la taille	GV
4	2-4-0	拿钥匙	na2yao4shi0	prendre la clé	GV
5	4-1-0	放风筝	fang4feng1zheng0	faire voler un cerf-volant	GV
6	4-2-0	盖房子	gai4fang2zi0	construire une maison	GV
7	4-4-0	做记号	zuo4ji4hao0	faire une marque	GV
8	3-2-0	买橘子	mai3ju2zi0	acheter des clémentines	GV
9	3-4-0	舍不得	she3bu4de0	ne pas abandonner	expression figée
10	1-1-1	出租车	chu1zu1che1	taxi	GN
11	2-1-1	直升机	zhi2sheng1ji1	hélicoptère	GN
12	2-2-1	床头灯	chuang2tou2deng1	lampe de chevet	GN
13	2-3-1	苹果汁	ping2guo3zhi1	jus de pomme	GN
14	2-4-1	博士生	bo2shi4sheng1	doctorant	GN
15	4-1-1	四千三	si4qian1san1	4,300	chiffre
16	4-2-1	自行车	zi4xing2che1	vélo	GN
17	4-3-1	摄影师	she4ying3shi1	photographe	GN
18	4-4-1	四万七	si4wan4qi1	47,000	chiffre
19	3-1-1	洗衣机	xi3yi1ji1	machine à laver	GN
20	3-2-1	五十一	wu3shi2yi1	51	chiffre
21	3-4-1	理发师	li3fa4shi1	coiffeur	GN
22	1-1-2	微波炉	wei1bo2lu2	micro-onde	GN
23	1-2-2	飞行员	fei1xing2yuan2	pilote	GN
24	1-3-2	三角形	san1jiao3xing2	triangle	GN
25	1-4-2	公寓楼	gong1yu4lou2	immeuble	GN
26	2-1-2	棉拖鞋	mian2tuo1xie2	pantoufle chaude	GN
27	2-2-2	柠檬茶	ning2meng2cha2	thé au citron	GN
28	2-3-2	游泳池	you2yong3chi2	piscine	GN
29	2-4-2	邮递员	you2di4yuan2	facteur	GN
30	3-1-2	保温瓶	bao3wen1ping2	thermos	GN
31	3-2-2	保洁员	bao3jie2yuan2	nettoyeur	GN
32	3-4-2	眼镜盒	yan3jing4he2	étui à lunettes	GN
33	4-1-2	健身房	jian4shen1fang2	gymnase	GN



34	4-2-2	数学题	shu4xue2ti2	exercice de maths	GN
35	4-3-2	电子琴	dian4zi3qin2	piano électrique	GN
36	4-4-2	向日葵	xiang4ri4kui2	tournesol	GN
37	1-2-3	芭蕾舞	ba1lei2wu3	ballet	GN
38	1-4-3	开胃酒	lai1wei4jiu3	apéritif	GN
39	2-1-3	停车场	ting2che1chang3	parking	GN
40	2-2-3	银行卡	yin2hang2ka3	carte bancaire	GN
41	2-4-3	博物馆	bo2wu4guan3	musée	GN
42	3-1-3	总经理	zong3jing1li3	directeur général	GN
43	3-2-3	五十五	wu3shi2wu3	55	chiffre
44	3-4-3	笔记本	bi3bi4ben3	cahier de notes	GN
45	4-1-3	化妆品	hua4zhuang1pin3	cosmétiques	GN
46	4-2-3	矿泉水	kuang4quan2shui3	eau minérale	GN
47	4-4-3	电热毯	dian4re4tan3	couverture chauffante	GN
48	1-1-4	星期四	xing1qi1si4	jeudi	GN
49	1-2-4	吸尘器	xi1chen2qi4	aspirateur	GN
50	1-3-4	终点站	zhong1dian3zhan4	terminus	GN
51	1-4-4	新电视	xin1dian4shi4	nouveau téléviseur	GN
52	2-1-4	十一月	shi2yi1yue4	novembre	GN
53	2-2-4	床头柜	chuang2tou2gui4	table de chevet	GN
54	2-3-4	皮手套	pi2shou3tao4	gants en cuir	GN
55	2-4-4	遥控器	yao2kong4qi4	télécommande	GN
56	3-1-4	火车票	huo3che1piao4	billet de train	GN
57	3-2-4	旅行社	lv3xing2she4	agence de voyage	GN
58	3-4-4	巧克力	qiao3ke4li4	chocolat	GN
59	4-1-4	面包店	mian4bao1dian4	boulangerie	GN
60	4-2-4	四十万	si4shi2wan4	400,000	chiffre
61	4-3-4	望远镜	wang4yuan3jing4	jumelles	GN
62	4-4-4	会议室	hui4yi4shi4	salle de réunion	GN

#### Quatre syllabes : 42 énoncés

N°	Ton	Phrase en chinois	Transcription pinyin	Traduction française	Commentaire
1	1-1-1-0	他搬家了	ta1ban1jia1le0	Il a déménagé.	SN(1)+SV(3)
2		二哥说的	er1ge1shuo1de0	Sange a dit cela.	SN(2)+SV(2)
3	2-1-1-0	桥开通了	qiao2kai1tong1le0	Le pont est ouvert.	SN(1)+SV(3)
4		明天休息	ming2tian1xiu1xi0	Demain on est en congé.	CCT(2) +V(2)
5	3-1-1-0	水烧开了	shui3shao1kai1le0	L'eau est bouillie.	SN(1)+SV(3)
6		火车开了	huo3che1kai1le0	Le train démarre.	SN(2)+SV(2)
7	4-1-1-0	路开通了	lu4kai1tong1le0	La route est ouverte.	SN(1)+SV(3)
8		大哥租的	da4ge1zu1de0	le grand frère l'a loué.	SN(2)+SV(2)
9	1-1-1-2	交通规则	jiao1tong1gui1ze2	Le code routier	SN(4)+SV(0)
10		张医生来	zhang1yi1sheng1lai2	Docteur Zhang viendra.	SN(3)+SV(1)
11		他刚说完	ta1gang1shuo1wan2	Il vient de terminer son discours.	SN(1)+SV(3)
12	2-1-1-2	长沙中学	chang2sha1zhong1xue2	le lycée de Changsha	SN(4)+SV(0)
13		王医生来	wang2yi1sheng1lai2	Docteur Wang vient.	SN(3)+SV(1)
14		桥刚修完	qiao2gang1xiu1wan2	Le pont vient d'être construit.	SN(1)+SV(3)

15	3-1-1-2	北京中学	bei3jing1zhong1xue2	le lycée de Beijing	SN(4)+SV(0)
16		李医生来	li3yi1sheng1lai2	Docteur Li vient.	SN(3)+SV(1)
17		你跟她来	ni3gen1talai2	Tu viens avec elle.	SN(1)+SV(3)
18	4-1-1-2	四双新鞋	si4shuang1xin1xie2	quatre paires de nouvelles chaussures	SN(4)+SV(0)
19		赵医生来	zhao4yi1sheng1lai2	Docteur Zhao vient	SN(3)+SV(1)
20		路刚修完	lu4gang1xiu1wan2	La route vient d'être construite.	SN(1)+SV(3)
21	1-1-1-3	三支铅笔	san1zhi1qian1bi3	trois crayons	SN(4)+SV(0)
22		方医生跑	fang1yi1sheng1pao3	Docteur Fang court.	SN(3)+SV(1)
23		车刚修好	che1gang1xiu1hao3	La voiture vient d'être réparée.	SN(1)+SV(3)
24	2-1-1-3	十只鹦鹉	shi2zhi1ying1wu3	dix perroquets	SN(4)+SV(0)
25		王医生跑	wang2yi1sheng1pao3	Docteur Wang court.	SN(3)+SV(1)
26		桥刚修好	qiao2gang1xiu1hao3	Le pont vient d'être construit.	SN(1)+SV(3)
27	3-1-1-3	五千一百	wu3qian1yi1bai3	5100	SN(4)+SV(0)
28		李医生跑	li3yi1sheng1pao3	Docteur Li court.	SN(3)+SV(1)
29		你跟他走	ni3gen1talzou3	Tu le suis.	SN(1)+SV(3)
30	4-1-1-3	四千三百	si4qian1yi1bai3	4300	SN(4)+SV(0)
31		赵医生跑	zhao4yi1sheng1pao3	Docteur Zhao court.	SN(3)+SV(1)
32		路刚修好	lu4gang1xiu1hao3	La route vient d'être construite.	SN(1)+SV(3)
33	1-1-1-4	公交车站	gong1jiao1che1zhan4	l'arrêt de l'autobus	SN(4)+SV(0)
34		张医生唱	zhang1yi1sheng1chang4	Docteur Zhang chante.	SN(3)+SV(1)
35		他刚出去	ta1gang1chulqu4	Il vient de sortir.	SN(1)+SV(3)
36	2-1-1-4	环山公路	huan2shan1gong1lu4	la route de montagne	SN(4)+SV(0)
37		王医生唱	wang2yi1sheng1chang4	Docteur Wang chante.	SN(3)+SV(1)
38	3-1-1-4	五千只鹤	wu3qian1zhi1he4	5000 grues	SN(4)+SV(0)
39		李医生唱	li4yi1sheng1chang4	Docteur Li chante.	SN(3)+SV(1)
40		你跟她去	ni3gen1talqu4	Tu y vas avec elle.	SN(1)+SV(3)
41	4-1-1-4	四天三夜	si4tian1san1ye4	quatre jours et trois nuits	SN(4)+SV(0)
42		赵医生唱	zhao4yi1sheng1chang4	Docteur Zhao chante.	SN(3)+SV(1)

### Neuf syllabes : 20 énoncés

N°	Ton	Phrase en chinois	Transcription pinyin	Traduction française	Commentaire
1	1-1-1-1-1-1-1-1-0	他喝西瓜汁跟咖啡了	ta1he1xi1gualzhi1gen1ka1fei1le0	Il a bu du jus de pastèque et du café.	SN + SV+ SN (1) (1) (6)
2		初中生跟高中生休息	chu1zhong1sheng1gen1gao1zhong1sheng1xiu1xi0	Les collégiens et les lycéens sont en congé.	SN(7) + SV(2)
3		他跟姨妈吃新鲜葡萄	ta1gen1yi1ma1chi1xin1xian1pu1tao0	Lui et la tante maternelle mangent des raisins frais.	SN + SV+ SN (4) (1) (4)
4		新司机偷偷开新车了	xin1si1ji1tou1tou1kai1xin1che2le0	Le nouveau chauffeur a secrètement conduit le nouveau véhicule.	SN +SV+ SN (3) (3) (2)
5	2-1-1-	姨妈喝西瓜汁跟花茶	yi2ma1he1xi1gualzhi1gen1hua1cha2	La tante maternelle boit du jus de pastèque et du thé de fleur.	SN + SV+ SN (2) (1) (6)

6	1-1-1-1-1-2	王医生他三姑妈出钱	wang2yi1sheng1ta1san1gulma1chulqian2	La troisième tante paternelle de Docteur Wang paie.	SN + SV+ SN (7) (1) (1)
7		姨妈跟她喝新鲜花茶	yi2ma1gen1talhelxin1xian1hua1cha2	La tante maternelle et elle prennent du thé de fleur frais.	SN + SV+ SN (4) (1) (4)
8		王医生帮她搬微波炉	wang2yi1sheng1bang1talban1wei1bolu2	Docteur Wang porte pour elle le micro-onde.	SN +SV+ SN (3) (3) (3)
9	4-1-1-1-1-1-1-2	四哥搬冰箱跟微波炉	si4ge1ban1bing1xiang1gen1wei1bolu2	Sige porte le frigo et le micro-onde.	SN + SV+ SN (2) (1) (6)
10		赵医生他三姑妈出钱	zhao4yi1sheng1ta1san1gulma1chulqian2	La troisième tante paternelle de Docteur Zhao paie.	SN + SV+ SN (7) (1) (1)
11		四哥跟她喝新鲜花茶	si4ge1gen1talhelxin1xian1hua1cha2	Sige et elle boivent du thé de fleur frais.	SN + SV+ SN (4) (1) (4)
12		赵医生帮她搬微波炉	zhao4yi1sheng1bang1talban1wei1bolu2	Docteur Zhao porte pour elle le micro-onde.	SN + SV+ SN (3) (3) (3)
13	2-1-1-1-1-1-1-4	刘哥搬冰箱跟新书架	liu2ge1ban1bing1xiang1gen1xin1shuljia4	Liuge porte le frigo et la nouvelle bibliothèque.	SN + SV+ SN (2) (1) (6)
14		王医生他三姑妈休假	wang2yi1sheng1ta1san1gulma1xiuljia4	La troisième tante paternelle de Docteur Wang prend un congé.	SN + SV+ SN (7) (1) (1)
15		姨妈跟她吃新鲜蔬菜	yi2ma1gen1talchilxin1xian1shulcai4	La tante et elle mangent des légumes frais.	SN + SV+ SN (4) (1) (4)
16		王医生帮她搬新书架	wang2yi1sheng1bang1talban1xin1shuljia4	Docteur Wang porte pour elle la nouvelle bibliothèque.	SN + SV+ SN (3) (3) (3)
17	4-1-1-1-1-1-1-4	四哥搬冰箱跟新书架	si4ge1ban1bing1xiang1gen1xin1shuljia4	Sige porte le frigo et la nouvelle bibliothèque.	SN + SV+ SN (2) (1) (6)
18		赵医生他三姑妈休假	zhao4yi1sheng1ta1san1gulma1xiuljia4	La troisième tante paternelle de Docteur Zhao prend un congé.	SN + SV+ SN (7) (1) (1)
19		四哥跟她吃新鲜茼蒿	si4ge1gen1talchilxin1xian1wolju4	Sige et elle mangent des laitues fraîches.	SN + SV+ SN (4) (1) (4)
20		赵医生帮她搬新书架	zhao4yi1sheng1bang1talban1xin1shuljia4	Docteur Zhao porte pour elle la nouvelle bibliothèque.	SN + SV+ SN (3) (3) (3)

## Annexe 2 : Interfaces des tests perceptifs en chinois mandarin et en français

Interfaces des tests perceptifs de la prosodie des affects sociaux chinois en chinois et en français

Test (2) \*

请通过移动下列条柱状按钮中的某一个来选择您感受到的态度，然后点击“回车”。

	陈述	
询问	失望	自信
怀疑-不轻信	蔑视	确信
赞赏	嘲讽	权威
礼貌	中性的惊讶	发怒
引诱	积极的惊讶	亲密-熟悉
妥协	消极的惊讶	母性

Test (2) \*

Sélectionnez l'attitude que vous avez perçue, puis cliquez sur la touche « entrée »

<input type="checkbox"/> DECLARATION		
<input type="checkbox"/> QUESTION	<input type="checkbox"/> DECEPTION	<input type="checkbox"/> CONFIANCE
<input type="checkbox"/> DOUTE-INCREDULITE	<input type="checkbox"/> MEPRIS	<input type="checkbox"/> EVIDENCE
<input type="checkbox"/> ADMIRATION	<input type="checkbox"/> IRONIE	<input type="checkbox"/> AUTORITE
<input type="checkbox"/> POLITESSE	<input type="checkbox"/> SURPRISE NEUTRE	<input type="checkbox"/> IRRITATION
<input type="checkbox"/> SEDUCTION	<input type="checkbox"/> SURPRISE POSITIVE	<input type="checkbox"/> FAMILIARITE
<input type="checkbox"/> RESIGNATION	<input type="checkbox"/> SURPRISE NEGATIVE	<input type="checkbox"/> LANGAGE MATERNEL

## Interfaces du test conceptuel des affects sociaux en chinois et en français

Perception\_Test (2) \*

妥协 询问

≠ ○ ○ ○ ○ ○ ○ ○ ○ ○ =

Perception\_Test (2) \*

QUESTION DÉCEPTION

≠ ○ ○ ○ ○ ○ ○ ○ ○ ○ =

## Interfaces des tests perceptifs des affects sociaux multimodaux en chinois et en français

Test (2) \*

您选择的 attitude 是

<input type="radio"/> 权威	<input type="radio"/> 陈述	<input type="radio"/> 确定
<input type="radio"/> 蔑视	<input type="radio"/> 儿向语	<input type="radio"/> 礼貌
<input type="radio"/> 失望	<input checked="" type="radio"/> 发怒	<input type="radio"/> 疑问
<input type="radio"/> 怀疑	<input type="radio"/> 惊讶	

Test (2) \*

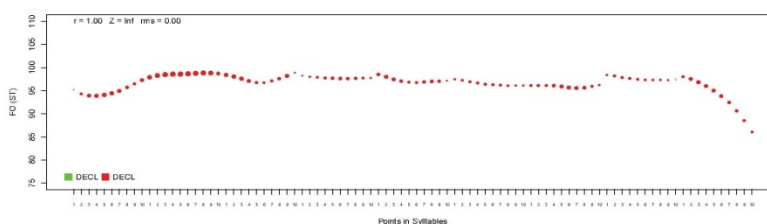
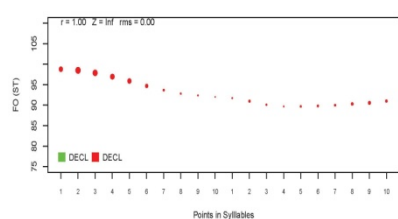
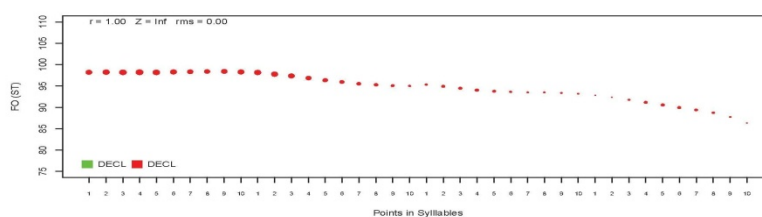
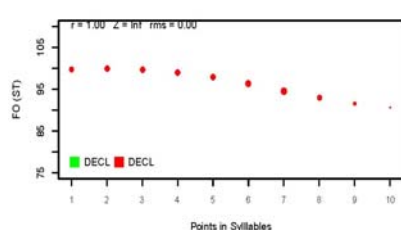
Quelle attitude avez-vous reconnue ?

<input type="radio"/> Autorité	<input type="radio"/> Déclaration	<input type="radio"/> Evidence
<input type="radio"/> Mépris	<input type="radio"/> Motherese	<input type="radio"/> Politesse
<input type="radio"/> Déception	<input type="radio"/> Irritation	<input type="radio"/> Question
<input type="radio"/> Doute	<input type="radio"/> Surprise neutre	

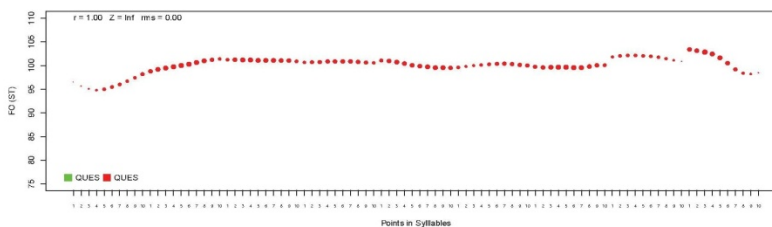
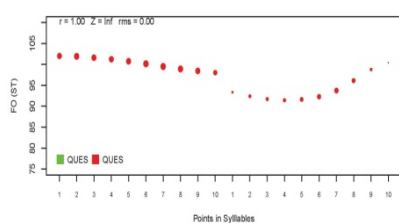
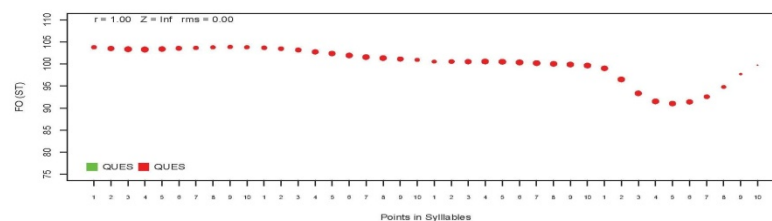
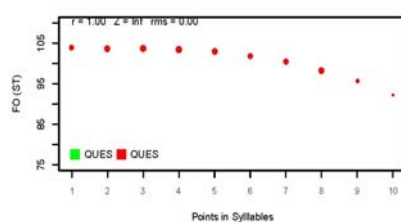
# Annexe 3 : Contours intonatifs des affects sociaux chinois

Le contour intonatif de chaque affect social pour chacune des longueurs

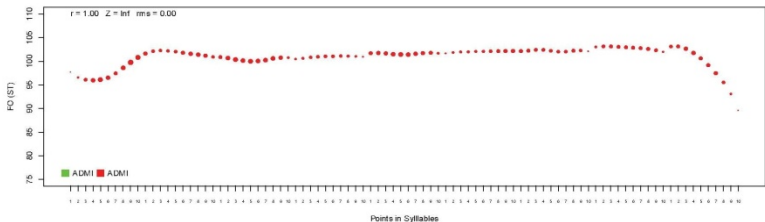
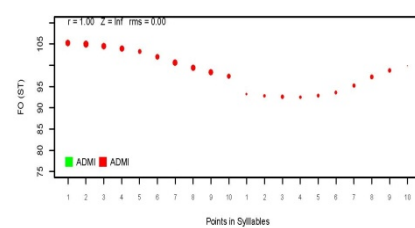
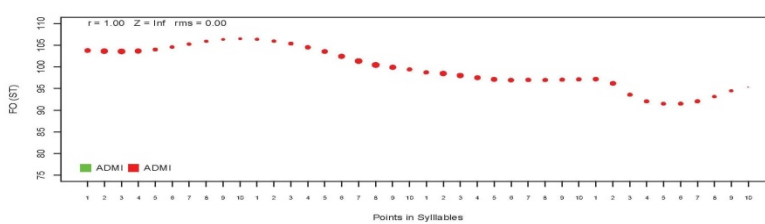
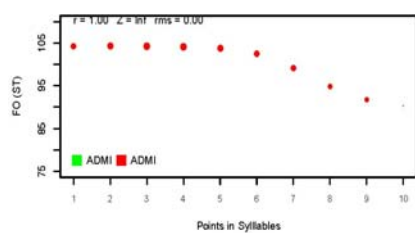
## Déclaration (DECL)



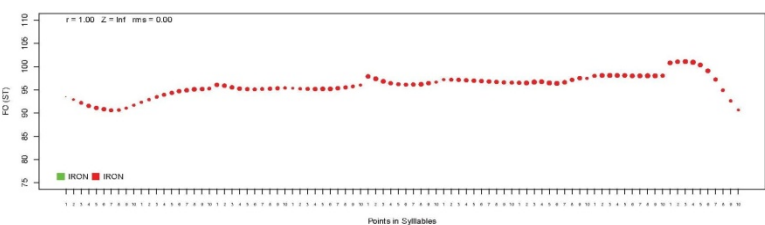
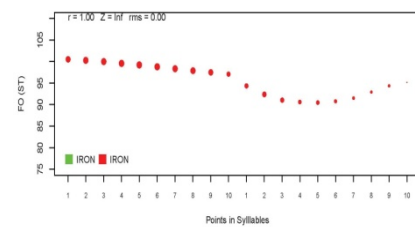
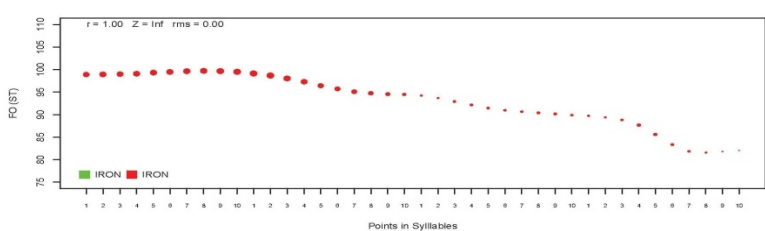
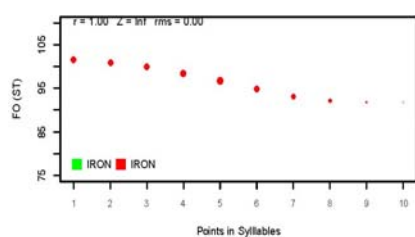
## Question (QUES)



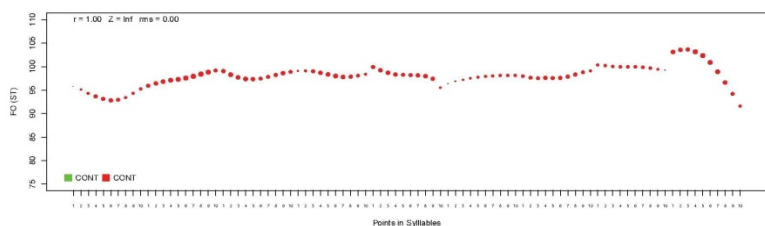
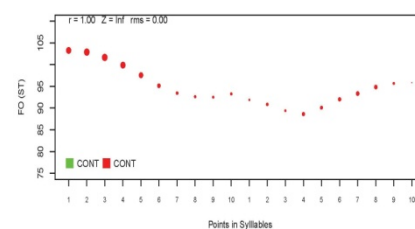
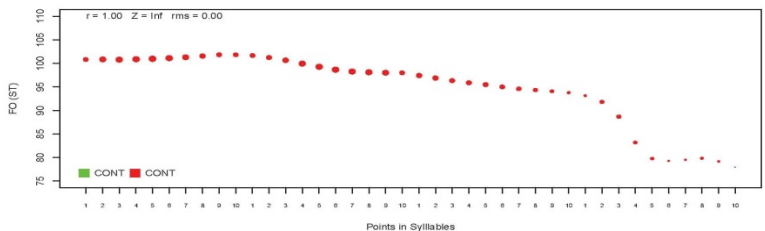
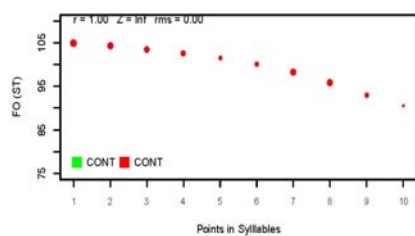
## Admiration (ADMI)



## Ironie (IRON)

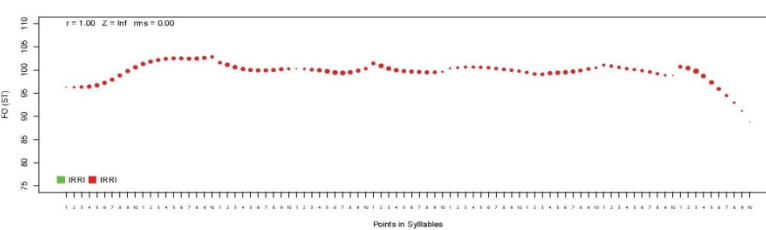
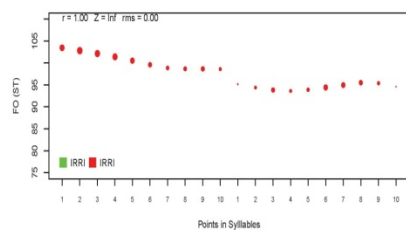
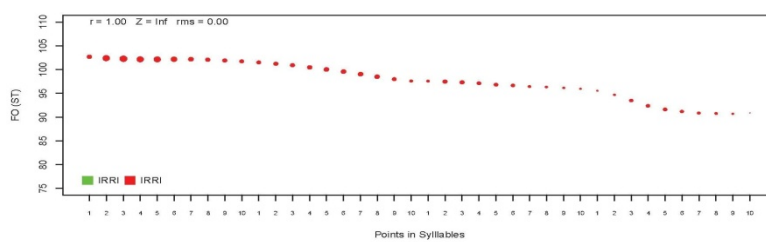
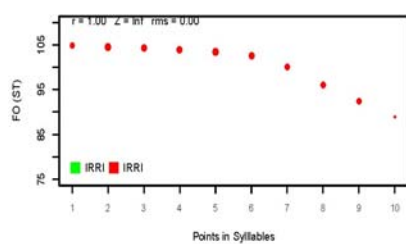


## Mépris (CONT)

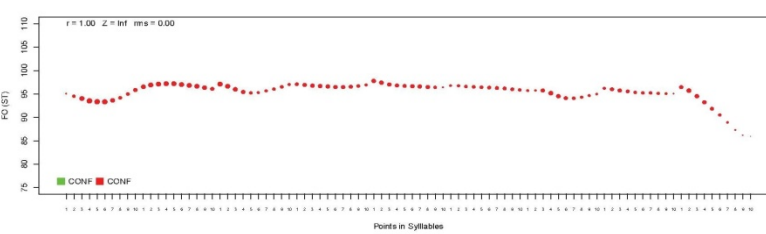
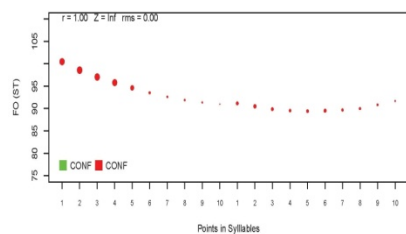
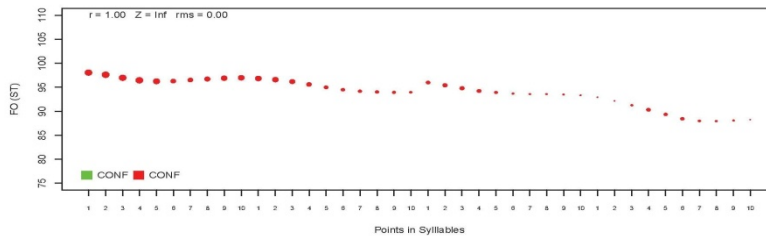
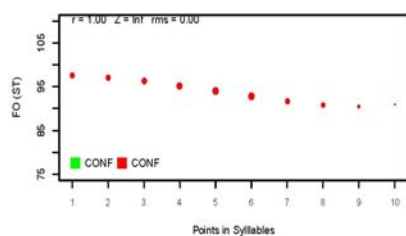




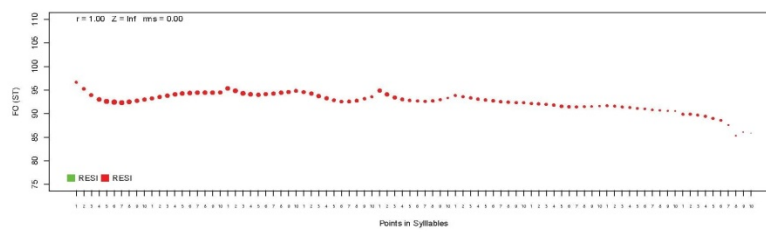
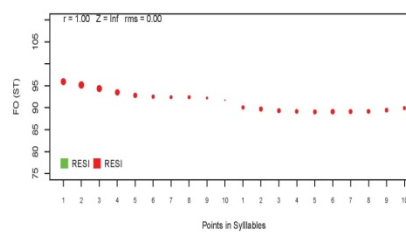
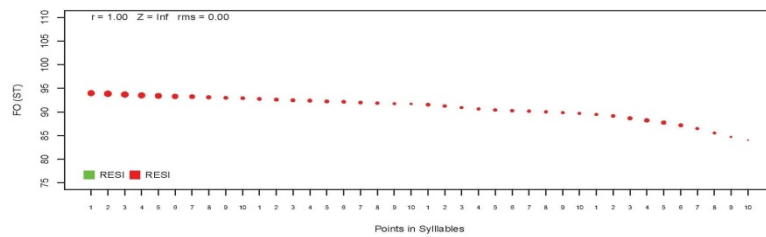
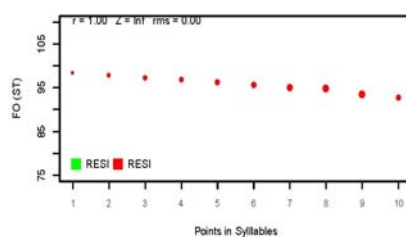
## Irritation (IRRI)



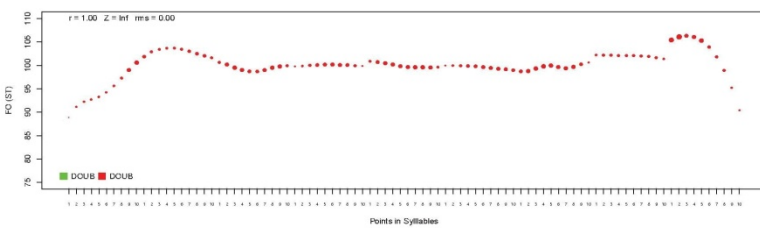
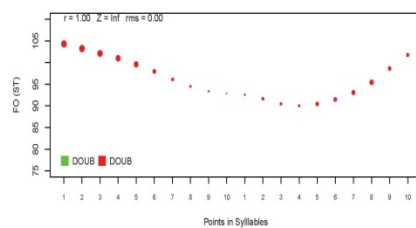
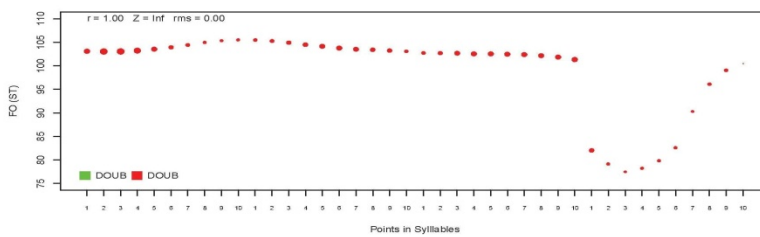
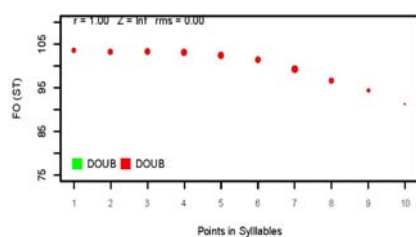
## Confiance (CONF)



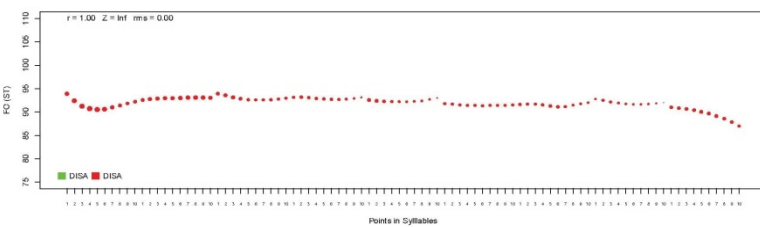
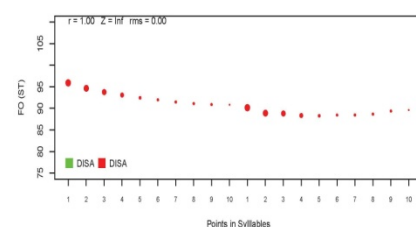
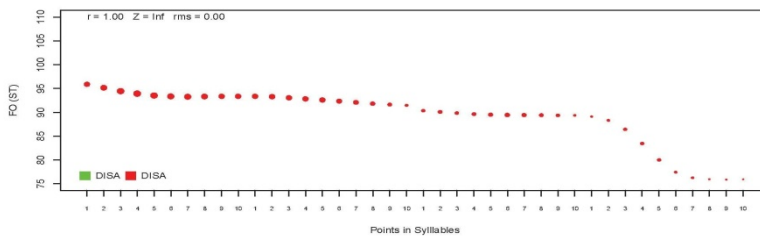
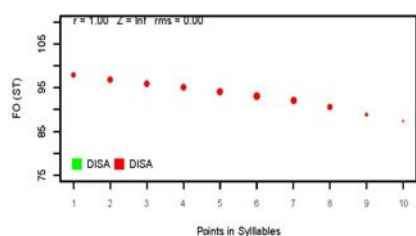
## Résignation (RESI)



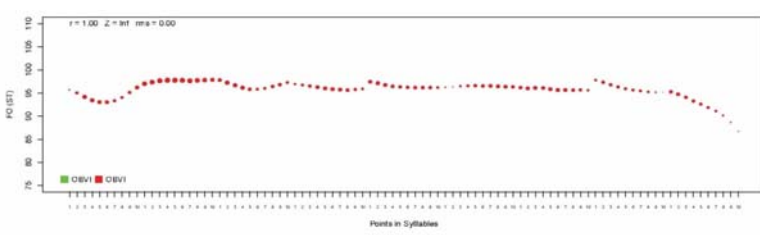
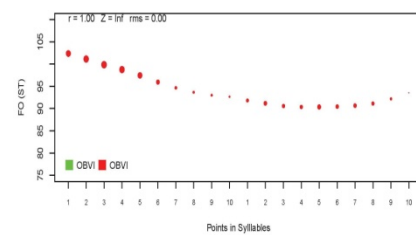
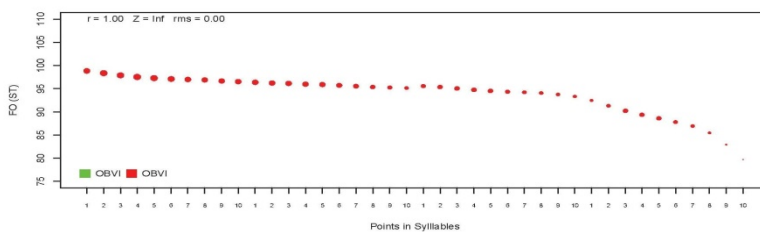
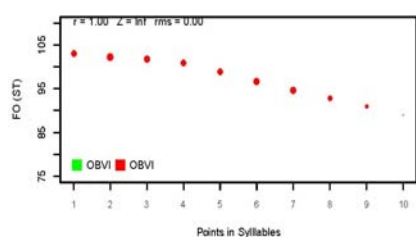
## Doute (DOUB)



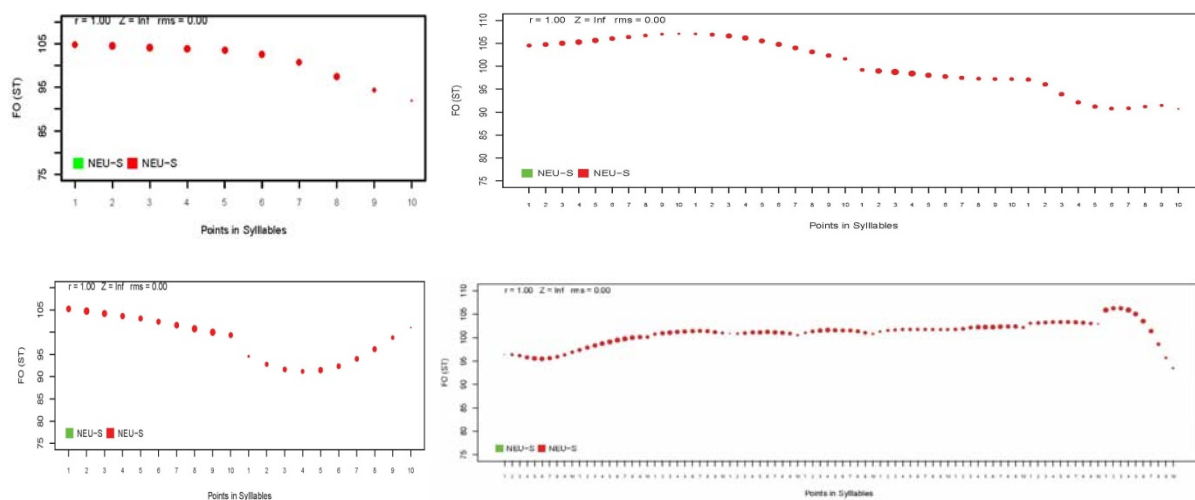
## Déception (DISA)



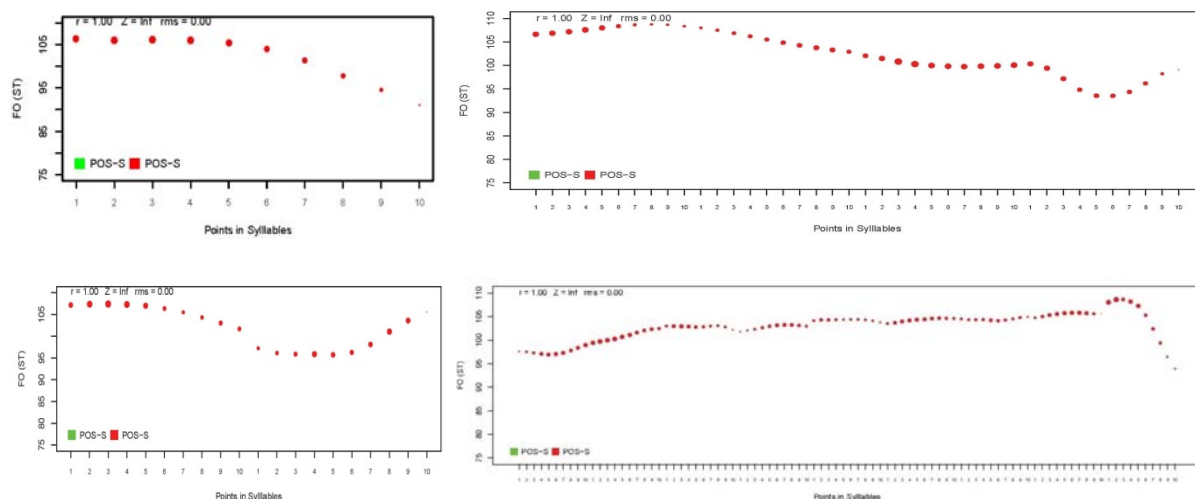
## Evidence (OBVI)



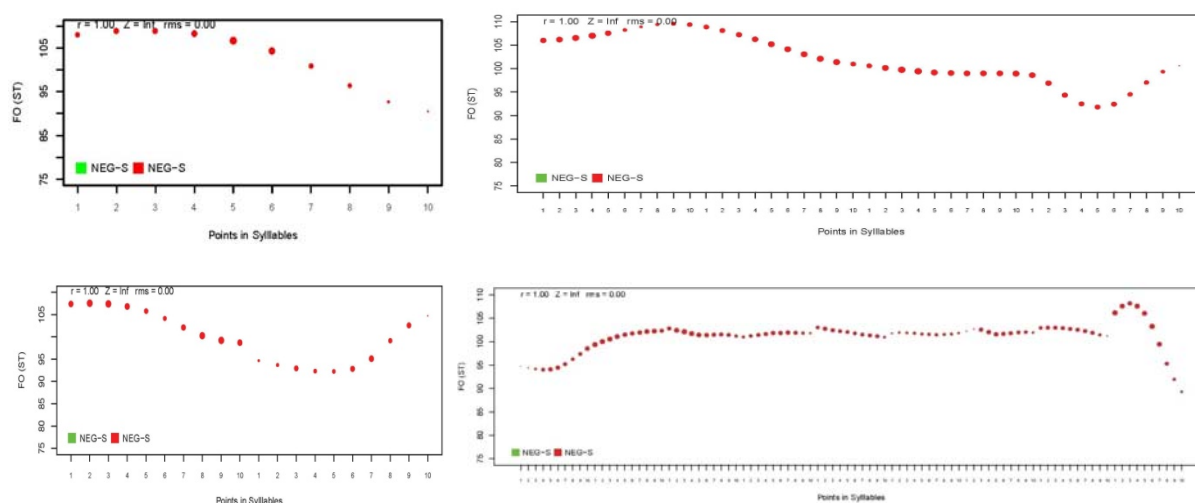
## Exclamation de surprise neutre (NEU-S)



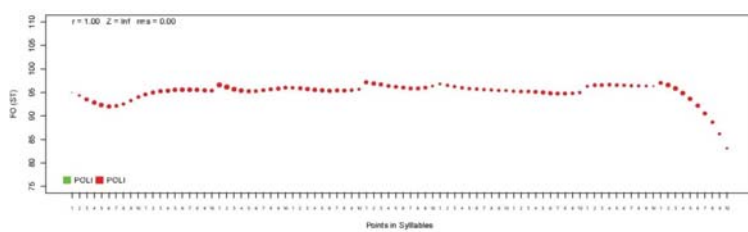
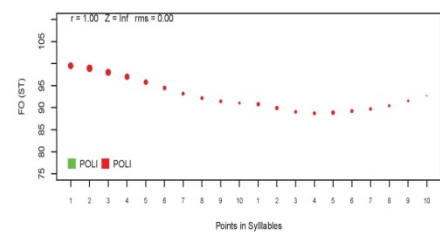
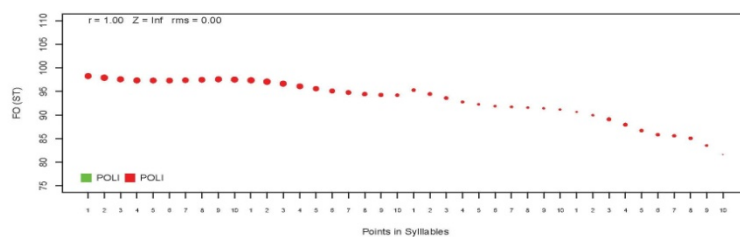
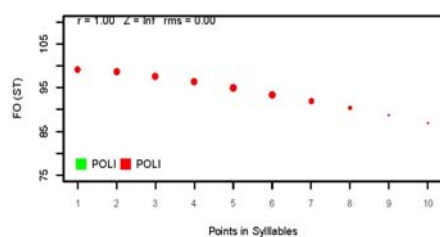
## Exclamation de surprise positive (POS-S)



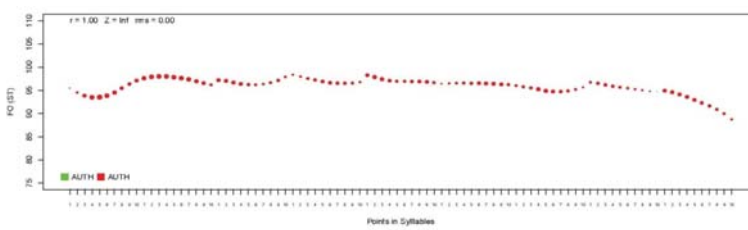
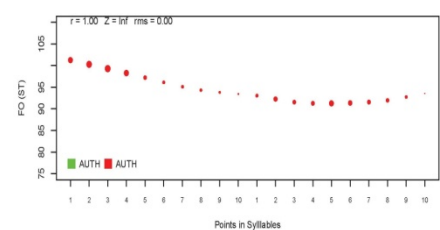
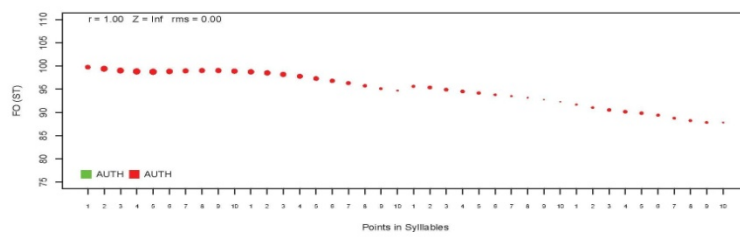
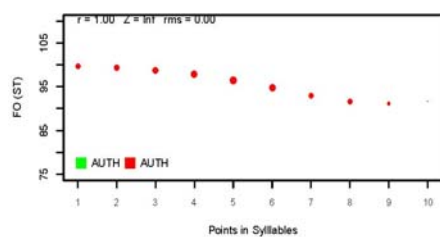
## Exclamation de surprise négative (NEG-S)



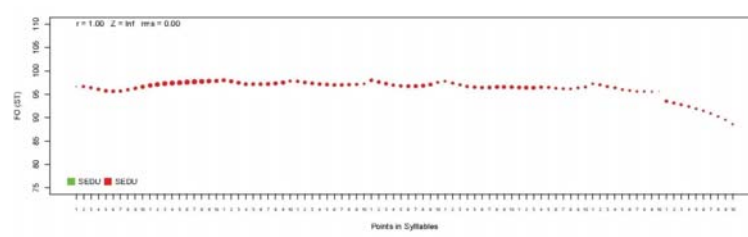
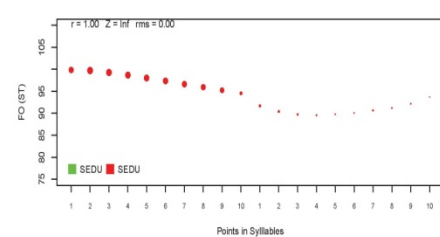
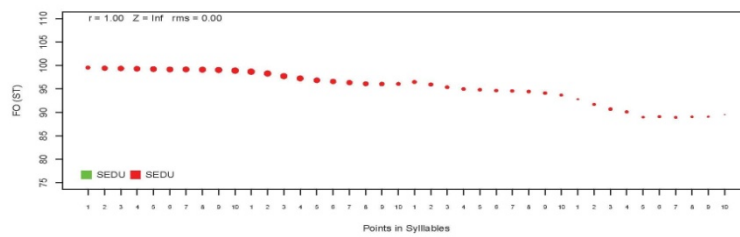
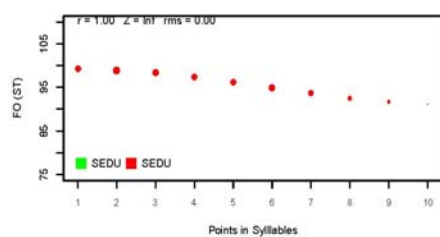
## Politesse (POLI)



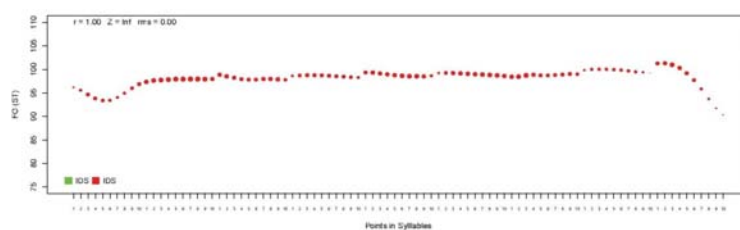
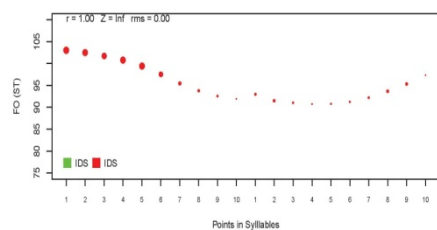
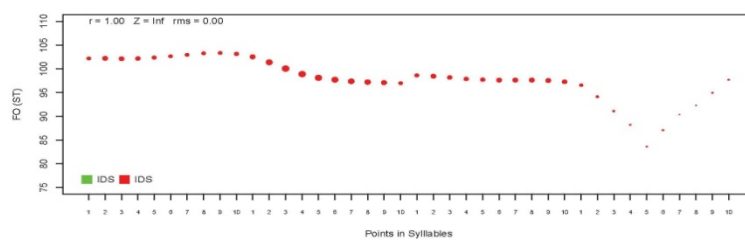
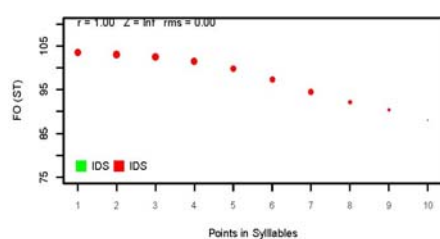
## Autorité (AUTH)



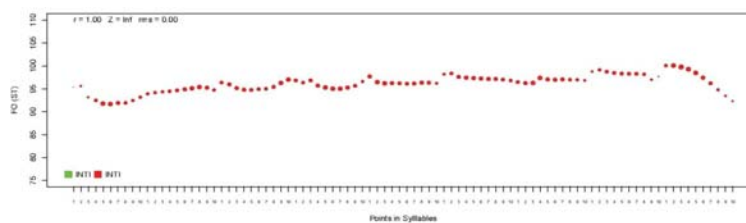
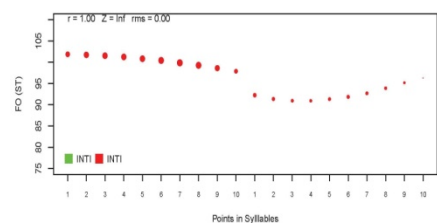
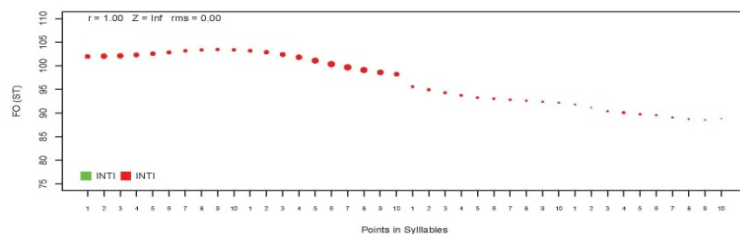
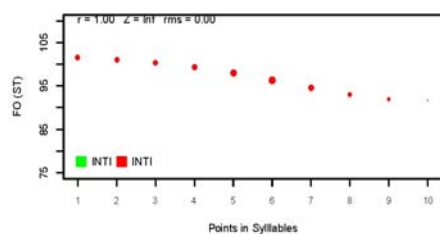
## Séduction (SEDU)



## Maternel (IDS)



## Intimité (INTI)



# Annexe 4 : Matrices de confusion des tests perceptifs

## 1. Matrice de confusion du test perceptif de la prosodie socio-affective chinoise par des sujets chinois (partie 2.3)

	ADMI	AUTH	CONF	DECL	DISA	DOUB	IRRI	IRON	INTI	CONT	IDS	OBVI	POLI	QUES	RESI	SEDU	NEG_S	NEU_S	POS_S
ADMI	11,92	1,91	3,34	11,92	1,59	8,43	1,59	2,23	6,04	3,02	3,5	7	5,88	2,38	0,64	1,43	4,61	17,17	5,41
AUTH	0,16	29,68	6,03	18,25	4,92	0,48	17,46	1,59	1,43	5,08	0,48	11,59	0,95	0,32	1,11	0	0,32	0,16	0
CONF	1,9	8,25	3,49	56,03	3,97	0,32	0,79	1,27	0,95	0,95	3,33	7,78	6,83	0,32	3,49	0	0	0,32	0
DECL	1,9	5,87	5,4	58,41	3,81	0,79	0,48	0,63	1,75	0,48	0,79	11,75	5,4	0,32	1,11	0,16	0,16	0,79	0
DISA	0,16	1,27	0,16	13,81	57,94	0,79	0,63	0,16	1,43	0,95	0,32	1,43	2,06	0,32	16,98	0,79	0,79	0	0
DOUB	1,59	2,23	1,75	6,04	0,32	35,14	0,48	6,52	2,86	9,86	0,48	2,38	0,79	11,61	0,48	0,64	4,13	11,76	0,95
IRRI	1,9	18,1	9,37	9,52	0,79	0,16	37,94	1,59	3,65	1,43	0,16	10,79	1,11	0,32	0,16	0,32	0,79	1,43	0,48
IRON	0,16	2,22	3,02	14,6	8,73	5,71	0,95	9,21	4,44	32,86	0,79	2,38	2,7	3,02	5,4	0,32	1,43	2,06	0
INTI	7,3	2,22	6,67	20	2,06	2,86	0,79	3,65	18,1	7,14	2,54	7,94	7,62	4,76	1,27	1,59	0,16	2,54	0,79
CONT	1,43	2,54	2,38	14,6	3,02	9,21	1,27	16,51	3,81	23,65	1,11	6,51	2,38	4,13	2,06	0,63	2,06	2,38	0,32
IDS	2,38	0,16	1,9	8,25	0,32	0,63	0	0,48	7,62	0,95	37,62	1,43	8,41	0,32	0,48	28,25	0	0,48	0,32
OBVI	0,95	11,76	7	38	7,15	1,91	4,93	1,27	2,23	3,34	0,48	12,4	4,29	0,48	2,38	0,16	0,64	0,64	0
POLI	2,7	1,11	2,86	49,21	1,43	0,32	0	0	5,71	0,16	10	2,06	20,16	0,16	2,38	1,75	0	0	0
QUES	3,49	1,9	5,4	10,48	0,16	18,25	0,95	1,9	4,92	2,06	0,48	3,97	2,38	28,57	0	0,79	0,79	10,95	2,54
RESI	0,16	1,43	1,11	21,62	43,56	0,16	0,48	0,95	2,23	2,07	0	3,18	1,43	0,32	20,35	0,48	0,32	0,16	0
SEDU	3,02	0,32	1,27	18,41	0,79	0	0	0	9,05	0,32	16,67	0,48	17,78	0	1,59	30	0	0,32	0
NEG_S	1,75	2,7	1,43	1,75	0,63	25,4	5,24	4,44	1,59	4,76	0,16	3,97	0,32	5,24	0	0,63	13,49	19,52	6,98
NEU_S	4,13	4,45	6,36	6,84	0,64	17,65	1,11	3,18	5,09	4,13	0,48	9,54	1,91	9,7	0,16	0,79	2,86	16,38	4,61
POS_S	7,46	0,95	2,38	2,38	0,79	14,76	1,11	1,43	5,56	0,79	1,59	2,54	2,22	4,92	0	3,17	6,19	14,76	26,98

## 2. Matrices de confusion du test perceptif de la prosodie socio-affective chinoise par des sujets français et vietnamiens (partie 3.2)

### Sujets français

	ADMI	AUTH	CONF	CONT	DECL	DISA	DOUB	IDS	INTI	IRONY	IRRI	NEG-S	NEU-S	OBVI	POLI	POS-S	QUES	RESI	SEDU
ADMI	10,42	2,92	4,17	2,5	3,75	4,58	7,5	6,67	2,5	5	3,75	7,5	7,5	7,92	2,92	4,17	7,5	0,83	7,92
AUTH	0	32,92	3,33	12,5	10	1,25	0,83	0,42	3,75	1,25	11,67	0,83	0,83	14,17	2,92	0,42	0,83	1,67	0,42
CONF	1,25	9,17	3,33	3,75	32,5	6,25	2,92	2,08	5	0,83	2,5	2,08	2,08	10,42	3,33	0,42	2,5	8,75	0,83
CONT	6,25	2,92	10,4	3,75	9,58	2,92	6,25	1,25	5	8,75	3,75	2,5	3,75	13,33	4,58	1,67	9,17	2,92	1,25
DECL	0,42	12,08	6,25	8,33	33,75	2,5	2,5	0,42	4,17	2,08	2,08	2,5	1,67	13,75	2,92	0,83	2,08	1,67	0
DISA	1,25	0,42	1,25	5	4,58	27,92	1,67	4,58	6,25	0,42	2,08	1,25	0,83	4,17	0,83	0,42	0	29,17	7,92
DOUB	1,25	1,25	4,58	2,5	5	7,5	17,5	1,25	5	5,42	3,33	2,92	11,25	7,92	3,33	2,08	15	2,08	0,83
IDS	5,83	0,42	2,92	2,5	5,42	3,75	4,58	27,5	4,58	4,17	0,42	1,25	2,5	2,5	5,83	1,25	2,08	4,17	18,33
INTI	0,42	5	7,08	0,83	20,83	1,25	4,17	1,25	7,5	2,08	3,33	1,67	4,17	15,42	4,58	0,83	17,5	2,08	0
IRON	2,08	3,75	5,42	4,17	27,08	5,83	2,92	2,5	10,4	2,08	2,5	1,25	3,75	5,83	3,33	0,42	10	4,58	2,08
IRRI	0	33,75	3,33	5,83	6,25	2,08	2,08	0,42	6,25	2,08	17,08	1,25	2,08	6,25	0,83	3,33	4,58	2,08	0,42
NEG_S	2,5	2,92	0,83	2,5	0,42	2,5	12,08	0	1,67	5,42	13,75	11,67	12,5	4,17	0,83	13,75	10,83	0,42	1,25
NEU_S	2,92	2,92	3,33	2,08	2,08	2,5	12,08	0,83	5	3,33	4,58	6,25	12,5	9,17	0,83	4,58	23,33	1,67	0
OBVI	0,83	14,58	4,58	4,17	30,42	4,58	2,5	2,5	5,42	1,25	2,08	1,67	1,67	12,92	3,33	0,42	2,5	4,17	0,42
POLI	2,08	1,25	10,8	2,08	22,92	5,42	4,17	6,25	7,08	0,83	0,83	1,25	2,08	4,58	10,42	0	1,67	7,08	9,17
POS_S	7,08	0,83	0,42	2,08	1,25	1,67	12,08	0,83	2,5	1,67	3,33	10,42	13,33	3,75	0,83	21,25	13,75	1,25	1,67
QUES	2,08	4,17	4,17	3,33	8,75	1,67	7,5	0	6,25	2,92	2,92	3,75	9,17	8,33	0,83	4,58	27,5	0,83	1,25
RESI	0,42	5	2,92	7,5	20	16,25	0,83	2,08	7,5	0,42	2,08	2,92	0,83	7,92	3,33	0	0,42	17,92	1,67
SEDU	5	0	4,17	0,83	5,42	7,5	1,67	17,92	4,58	1,67	0,42	1,25	1,25	1,25	10	0	1,25	3,75	32,08

### Sujets vietnamiens

	ADMI	AUTH	CONF	CONT	DECL	DISA	DOUB	IDS	INTI	IRONY	IRRI	NEG-S	NEU-S	OBVI	POLI	POS-S	QUES	RESI	SEDU
ADMI	10,83	3,33	4,17	4,17	3,75	1,67	7,92	4,58	3,33	6,67	2,50	5,42	9,17	4,58	4,58	7,08	7,08	2,92	6,25
AUTH	0,83	23,75	10,00	6,67	15,42	2,92	2,50	2,50	1,67	2,92	12,92	0,42	2,08	4,58	1,25	0	2,92	6,25	0,42
CONF	1,67	3,75	8,33	5,00	27,50	8,75	2,50	1,25	4,58	2,50	2,08	4,17	2,92	7,08	7,08	0	5,42	4,17	1,25
CONT	5,83	2,92	6,25	9,17	11,67	2,08	9,58	1,67	5,00	10,42	3,33	1,67	5,00	6,25	4,17	1,67	9,17	2,50	1,67
DECL	2,50	7,08	11,67	2,50	23,33	5,00	2,08	0,42	6,25	1,25	3,75	3,75	3,33	10,00	3,75	0,83	4,17	7,50	0,83
DISA	2,08	0	2,50	0	6,67	46,67	4,17	1,25	2,08	1,25	2,08	7,92	2,08	0,42	2,92	0	0,83	15,42	1,67
DOUB	4,17	2,92	5,83	4,58	6,25	2,50	10,83	0,83	3,33	6,25	2,08	2,92	10,83	7,50	5,42	3,33	18,75	1,25	0,42
IDS	5,83	0,42	4,17	5,42	5,42	0,42	0,83	28,33	3,75	5,00	0,42	1,25	1,25	4,17	12,50	0	0,83	1,67	18,33
INTI	2,50	5,83	10,83	1,67	12,08	1,67	2,92	1,25	7,50	4,58	4,17	0,83	5,83	14,17	5,42	2,92	12,50	2,50	0,83
IRON	1,67	2,92	7,50	5,83	20,00	4,17	5,00	0	7,50	3,33	3,75	3,75	6,67	10,00	4,58	0	6,25	3,75	3,33
IRRI	2,08	31,67	3,75	5,00	6,67	0,42	2,50	0,83	2,08	2,08	22,92	0,42	2,92	4,58	1,25	0	2,92	7,92	0
NEG_S	5,00	7,08	4,58	7,50	0,83	0,42	7,50	0,83	1,67	6,67	8,33	7,08	9,58	7,08	1,67	10,00	11,67	2,08	0,42
NEU_S	3,33	11,67	7,08	2,92	5,00	0,83	6,67	1,67	3,33	1,67	9,58	3,75	9,17	2,50	1,25	8,75	19,58	1,25	0
OBVI	1,25	7,50	12,08	5,83	21,67	7,08	4,58	1,25	5,00	0,42	4,58	0,83	4,17	10,42	2,92	0,42	3,75	3,75	2,50
POLI	4,17	1,25	3,33	0,42	30,00	9,58	2,50	2,08	6,67	1,67	0,83	4,17	3,75	4,58	15,00	0	0,42	6,67	2,92
POS_S	7,08	1,67	4,58	2,92	2,50	0,42	9,17	1,25	1,67	3,33	2,92	6,67	10,42	3,75	2,92	20,00	16,25	1,25	1,25
QUES	2,08	5,00	8,33	0,83	4,58	1,25	9,17	1,67	5,83	2,92	5,00	2,08	7,92	9,58	1,25	6,25	23,33	2,08	0,83
RESI	0,83	3,33	2,50	1,67	14,17	29,58	6,25	1,67	4,17	2,92	1,67	6,25	3,33	3,75	1,67	0	0,42	15,42	0,42
SEDU	3,75	0,42	2,92	0,42	11,67	7,92	3,75	9,17	10,00	0,83	1,25	1,67	1,25	2,50	15,42	0	0,83	4,58	21,67

### 3. Matrices de confusion du test perceptif des affects sociaux multimodaux du français par des sujets français (partie 4.4) et par des sujets chinois (partie 4.5)

#### Sujets français

##### Audio seul

	AUTH	CONT	DECL	DISA	DOUB	IDS	IRRI	NEU-S	OBVI	POLI	QUES
AUTH	64.29	7.14	0	0	0	0	28.57	0	0	0	0
CONT	0	28.57	21.43	14.29	0	7.14	21.43	0	7.14	0	0
DECL	0	0	92.86	0	0	0	0	0	0	7.14	0
DISA	0	14.29	0	78.57	0	0	7.14	0	0	0	0
DOUB	0	0	0	0	28.57	0	0	35.71	0	0	35.71
IDS	0	0	7.14	0	0	85.71	0	0	7.14	0	0
IRRI	28.57	0	0	0	0	0	71.43	0	0	0	0
NEU-S	0	0	0	0	21.43	0	0	64.29	0	0	14.29
OBVI	14.29	7.14	0	0	7.14	0	0	7.14	57.14	0	7.14
POLI	0	0	21.43	0	0	14.29	0	0	14.29	50	0
QUES	0	0	0	0	21.43	0	0	14.29	0	0	64.29

##### Vidéo seul

	AUTH	CONT	DECL	DISA	DOUB	IDS	IRRI	NEU-S	OBVI	POLI	QUES
AUTH	50	0	7.14	0	0	0	14.29	0	28.57	0	0
CONT	0	85.71	0	0	0	0	14.29	0	0	0	0
DECL	7.14	0	71.43	0	0	0	7.14	0	14.29	0	0
DISA	0	7.14	7.14	71.43	0	0	14.29	0	0	0	0
DOUB	0	0	0	14.29	57.14	0	7.14	7.14	0	0	14.29
IDS	7.14	0	21.43	0	0	21.43	0	0	21.43	28.57	0
IRRI	28.57	7.14	0	7.14	0	0	57.14	0	0	0	0
NEU-S	0	0	7.14	14.29	7.14	0	14.29	28.57	14.29	0	14.29
OBVI	21.43	0	14.29	0	0	0	7.14	14.29	42.86	0	0
POLI	0	0	21.43	0	0	14.29	0	0	7.14	57.14	0
QUES	14.29	0	57.14	7.14	0	0	0	14.29	7.14	0	0

##### Audio-visuel

	AUTH	CONT	DECL	DISA	DOUB	IDS	IRRI	NEU-S	OBVI	POLI	QUES
AUTH	78.57	0	0	0	0	0	14.29	0	7.14	0	0
CONT	0	92.86	0	0	0	0	7.14	0	0	0	0
DECL	0	0	100	0	0	0	0	0	0	0	0
DISA	0	7.14	0	85.71	0	0	7.14	0	0	0	0
DOUB	0	0	0	0	85.71	0	0	7.14	0	0	7.14
IDS	0	0	0	0	0	85.71	0	0	0	14.29	0
IRRI	21.43	7.14	0	0	0	0	71.43	0	0	0	0
NEU-S	0	0	0	0	21.43	0	0	42.86	0	0	35.71
OBVI	0	0	0	0	0	0	7.14	0	92.86	0	0
POLI	0	0	7.14	0	0	28.57	0	0	0	64.29	0
QUES	0	0	7.14	0	14.29	0	0	28.57	0	7.14	42.86



## Sujets chinois

### Audio seul

	AUTH	CONT	DECL	DISA	DOUB	IDS	IRRI	NEU-S	OBVI	POLI	QUES
AUTH	28.57	3.57	0	0	0	0	60.71	0	7.14	0	0
CONT	10.71	14.29	32.14	14.29	0	3.57	0	0	14.29	10.71	0
DECL	3.57	7.14	67.86	0	0	0	0	0	14.29	7.14	0
DISA	0	14.29	3.57	78.57	0	0	0	0	3.57	0	0
DOUB	0	3.57	0	3.57	39.29	0	0	17.86	0	0	35.71
IDS	0	0	0	0	0	85.71	0	0	3.57	7.14	3.57
IRRI	10.71	3.57	0	7.14	0	0	67.86	0	10.71	0	0
NEU-S	0	3.57	0	3.57	32.14	0	0	17.86	0	0	42.86
OBVI	10.71	10.71	0	0	28.57	3.57	0	28.57	10.71	0	7.14
POLI	0	0	28.57	0	0	21.43	0	0	3.57	46.43	0
QUES	0	7.14	0	0	7.14	0	0	0	0	14.29	71.43

### Vidéo seul

	AUTH	CONT	DECL	DISA	DOUB	IDS	IRRI	NEU-S	OBVI	POLI	QUES
AUTH	35.71	0	0	3.57	0	0	17.86	0	42.86	0	0
CONT	0	57.14	0	10.71	7.14	0	17.86	7.14	0	0	0
DECL	10.71	14.29	57.14	3.57	0	0	0	0	14.29	0	0
DISA	3.57	7.14	14.29	60.71	3.57	3.57	0	0	3.57	3.57	0
DOUB	3.57	0	0	21.43	39.29	0	0	7.14	3.57	0	25
IDS	7.14	0	7.14	0	0	42.86	0	0	7.14	35.71	0
IRRI	21.43	3.57	0	0	7.14	0	53.57	0	10.71	0	3.57
NEU-S	3.57	3.57	50	17.86	3.57	0	0	0	7.14	3.57	10.71
OBVI	7.14	7.14	0	7.14	0	0	21.43	3.57	53.57	0	0
POLI	3.57	0	7.14	0	0	25	0	0	0	64.29	0
QUES	21.43	7.14	53.57	7.14	0	0	0	0	7.14	3.57	0

### Audio-visuel

	AUTH	CONT	DECL	DISA	DOUB	IDS	IRRI	NEU-S	OBVI	POLI	QUES
AUTH	57.14	3.57	0	0	0	0	21.43	0	17.86	0	0
CONT	0	71.43	0	28.57	0	0	0	0	0	0	0
DECL	0	3.57	89.29	3.57	0	0	0	0	3.57	0	0
DISA	0	10.71	3.57	85.71	0	0	0	0	0	0	0
DOUB	0	3.57	0	0	71.43	0	0	7.14	0	0	17.86
IDS	0	0	0	0	0	89.29	0	0	0	10.71	0
IRRI	14.29	7.14	0	7.14	0	0	67.86	0	3.57	0	0
NEU-S	0	0	0	0	25	3.57	0	32.14	0	0	39.29
OBVI	14.29	14.29	3.57	3.57	0	0	0	10.71	50	0	3.57
POLI	0	0	3.57	0	0	25	0	0	3.57	67.86	0
QUES	0	3.57	35.71	3.57	10.71	0	0	0	0	0	46.43

#### 4. Matrices de confusion du test perceptif des affects sociaux multimodaux du chinois par des sujets chinois (partie 2.5) et par des sujets français (partie 4.6)

##### Sujets chinois

##### Audio seul

	AUTH	CONT	DECL	DISA	DOUB	IDS	IRRI	NEU-S	OBVI	POLI	QUES
AUTH	37.5	3.12	12.5	6.25	0	0	9.38	0	31.25	0	0
CONT	3.12	21.88	6.25	6.25	28.12	0	3.12	12.5	9.38	0	9.38
DECL	3.12	3.12	50	0	0	0	0	0	37.5	6.25	0
DISA	6.25	6.25	34.38	31.25	0	3.12	0	0	9.38	9.38	0
DOUB	0	12.5	9.38	6.25	21.88	0	3.12	12.5	9.38	0	25
IDS	0	3.12	12.5	3.12	9.38	46.88	0	3.12	6.25	9.38	6.25
IRRI	18.75	9.38	9.38	3.12	3.12	0	37.5	3.12	12.5	0	3.12
NEU-S	0	0	0	0	12.5	0	6.25	65.62	6.25	0	9.38
OBVI	15.62	6.25	21.88	3.12	0	0	0	0	46.88	3.12	3.12
POLI	15.62	0	56.25	0	0	3.12	0	0	6.25	18.75	0
QUES	0	3.12	6.25	3.12	18.75	3.12	0	12.5	0	0	53.12

##### Vidéo seul

	AUTH	CONT	DECL	DISA	DOUB	IDS	IRRI	NEU-S	OBVI	POLI	QUES
AUTH	25	15.62	21.88	6.25	0	3.12	0	0	25	0	3.12
CONT	0	15.62	9.38	31.25	12.5	0	3.12	3.12	9.38	9.38	6.25
DECL	6.25	0	53.12	3.12	3.12	3.12	0	0	18.75	12.5	0
DISA	0	25	6.25	62.5	0	0	0	0	0	0	6.25
DOUB	0	0	0	9.38	53.12	0	6.25	3.12	0	3.12	25
IDS	6.25	0	15.62	0	0	25	0	0	15.62	31.25	6.25
IRRI	0	18.75	0	31.25	6.25	0	40.62	0	0	0	3.12
NEU-S	3.12	0	3.12	6.25	18.75	6.25	0	18.75	6.25	6.25	31.25
OBVI	15.62	15.62	3.12	12.5	3.12	0	0	0	25	18.75	6.25
POLI	0	0	18.75	0	0	9.38	0	6.25	12.5	50	3.12
QUES	9.38	0	43.75	0	0	9.38	0	0	12.5	15.62	9.38

##### Audio-visuel

	AUTH	CONT	DECL	DISA	DOUB	IDS	IRRI	NEU-S	OBVI	POLI	QUES
AUTH	56.25	0	0	3.12	0	0	0	0	40.62	0	0
CONT	0	28.12	6.25	15.62	25	0	3.12	0	6.25	0	15.62
DECL	3.12	0	53.12	0	0	0	0	0	34.38	9.38	0
DISA	0	6.25	3.12	90.62	0	0	0	0	0	0	0
DOUB	0	6.25	0	0	46.88	0	6.25	9.38	0	0	31.25
IDS	0	3.12	9.38	0	3.12	59.38	0	0	9.38	15.62	0
IRRI	0	31.25	3.12	6.25	0	0	59.38	0	0	0	0
NEU-S	0	0	3.12	0	15.62	0	0	71.88	0	0	9.38
OBVI	15.62	6.25	25	0	0	0	0	0	46.88	6.25	0
POLI	0	0	34.38	0	0	9.38	0	0	9.38	46.88	0
QUES	0	0	6.25	0	25	3.12	0	15.62	3.12	3.12	43.75

## Sujets français

### Audio seul

	AUTH	CONT	DECL	DISA	DOUB	IDS	IRRI	NEU-S	OBVI	POLI	QUES
AUTH	35.71	7.14	21.43	0	0	0	17.86	0	17.86	0	0
CONT	3.57	14.29	7.14	7.14	17.86	3.57	7.14	0	10.71	3.57	25
DECL	14.29	3.57	50	0	0	0	0	0	17.86	14.29	0
DISA	0	7.14	7.14	57.14	0	3.57	7.14	3.57	7.14	7.14	0
DOUB	7.14	7.14	14.29	10.71	14.29	0	7.14	21.43	0	0	17.86
IDS	0	3.57	10.71	3.57	0	57.14	0	7.14	7.14	7.14	3.57
IRRI	39.29	7.14	0	3.57	3.57	0	28.57	7.14	10.71	0	0
NEU-S	0	7.14	0	3.57	7.14	0	28.57	14.29	21.43	0	17.86
OBVI	14.29	14.29	42.86	0	7.14	0	0	0	14.29	3.57	3.57
POLI	0	3.57	57.14	14.29	0	14.29	3.57	0	7.14	0	0
QUES	0	0	0	7.14	25	0	0	35.71	3.57	0	28.57

### Vidéo seul

	AUTH	CONT	DECL	DISA	DOUB	IDS	IRRI	NEU-S	OBVI	POLI	QUES
AUTH	39.29	7.14	14.29	7.14	0	0	3.57	0	25	3.57	0
CONT	0	17.86	3.57	28.57	21.43	0	25	0	0	0	3.57
DECL	0	3.57	64.29	10.71	3.57	0	0	3.57	7.14	7.14	0
DISA	0	7.14	3.57	53.57	17.86	0	3.57	7.14	7.14	0	0
DOUB	0	0	0	7.14	39.29	0	28.57	3.57	3.57	0	17.86
IDS	3.57	0	25	0	0	25	0	3.57	10.71	28.57	3.57
IRRI	0	25	0	17.86	7.14	0	50	0	0	0	0
NEU-S	7.14	0	10.71	25	3.57	0	7.14	7.14	14.29	0	25
OBVI	3.57	7.14	21.43	17.86	0	0	0	3.57	39.29	7.14	0
POLI	3.57	0	21.43	0	0	10.71	3.57	7.14	25	28.57	0
QUES	0	3.57	46.43	0	7.14	0	0	7.14	10.71	17.86	7.14

### Audio-visuel

	AUTH	CONT	DECL	DISA	DOUB	IDS	IRRI	NEU-S	OBVI	POLI	QUES
AUTH	64.29	7.14	7.14	0	0	0	10.71	0	10.71	0	0
CONT	0	21.43	7.14	46.43	3.57	0	14.29	0	0	0	7.14
DECL	3.57	0	75	0	0	0	0	0	7.14	14.29	0
DISA	0	3.57	0	85.71	0	3.57	3.57	3.57	0	0	0
DOUB	0	7.14	0	7.14	53.57	0	14.29	10.71	0	0	7.14
IDS	0	0	10.71	0	0	78.57	0	3.57	0	3.57	3.57
IRRI	3.57	32.14	0	3.57	7.14	0	53.57	0	0	0	0
NEU-S	3.57	10.71	0	0	3.57	0	28.57	32.14	7.14	0	14.29
OBVI	14.29	7.14	14.29	0	0	0	0	0	60.71	3.57	0
POLI	0	0	21.43	0	3.57	14.29	0	0	7.14	53.57	0
QUES	0	0	10.71	0	10.71	0	0	35.71	10.71	10.71	21.43